

## Rubrique : articles de synthèses / revues générales

**Titre. La RT-qPCR en oncologie : considérations pour la normalisation**

**Title. Considerations for normalisation of RT-qPCR in oncology**

Alexandre Ho-Pun-Cheung<sup>1,2,3</sup>

Dominic Cellier<sup>2</sup>

Evelyne Lopez-Crapez<sup>3</sup>

<sup>1</sup> INSERM U860, CRLC Val d'Aurelle, Montpellier

<sup>2</sup> Merck Santé, Lyon

<sup>3</sup> Laboratoire d'oncobiologie, CRLC Val d'Aurelle, Montpellier

Correspondance : Evelyne Lopez-Crapez, Centre de Recherche en Cancérologie, C.R.L.C. Val d'Aurelle, 208 rue des Apothicaires, 34298 Montpellier Cedex 5, France.

Tel: (+33) 467-613-048; Fax: (+33) 467-632-873

E-mail: [ecrapez@valdorel.fnclcc.fr](mailto:ecrapez@valdorel.fnclcc.fr)

**Résumé.** L'analyse de l'expression génique présente de multiples applications en cancérologie, notamment au niveau du diagnostic, du pronostic, et de la prise en charge thérapeutique. Dans ce domaine, la RT-qPCR (*reverse transcription quantitative polymerase chain reaction*) est devenue la méthode de référence pour la quantification des ARNm.

Néanmoins, c'est une technique délicate mettant en jeu différentes étapes incluant l'extraction de l'ARN, la synthèse d'ADNc, la PCR quantitative, et l'analyse, qui sont toutes sources de variation. L'obtention de résultats biologiques pertinents est dépendante de la mise place d'une normalisation par rapport à une référence, telle qu'une quantité étalon de cellules, d'ARN, ou d'un gène de ménage. Le choix de la méthode de normalisation est crucial, et toute variation propre à la référence va altérer les résultats normalisés. En passant en revue les différentes approches de normalisation, l'objectif de cet article est d'aborder les problèmes pouvant y être associés, afin d'en tirer des recommandations pour l'élaboration d'une stratégie appropriée à la RT-qPCR appliquée à l'oncologie.

**Mots clés :** *RT-PCR, quantification, normalisation, gène de référence, cancer*

**Abstract.** Gene expression analysis has many applications in the management of cancer, including diagnosis, prognosis, and therapeutic care. In this context, the reverse transcription quantitative polymerase chain reaction (RT-qPCR) has become the "gold standard" for mRNA quantification. However, this technique involves several critical steps such as RNA extraction, cDNA synthesis, quantitative PCR, and analysis, which all can be source of variation. To obtain biologically meaningful results, data normalisation is required to correct sample-to-sample variations that may be introduced during this multistage

process. Normalisation can be carried out against a housekeeping gene, total RNA mass, or cell number. Careful choice of the normalization method is crucial, as any variation in the reference will introduce errors in the quantification of mRNA transcripts. By reviewing the different methods available and their related problems, the aim of this article is to provide recommendations for the set up of an appropriate normalisation strategy for RT-qPCR data in oncology.

**Key words :** *RT-PCR, quantification, normalisation, housekeeping gene, cancer*

Les cellules tumorales montrent typiquement des dérégulations de l'expression de gènes responsables des caractéristiques néoplasiques telles que la prolifération incontrôlée, l'inhibition de l'apoptose, l'invasion, l'angiogenèse ou la formation de métastases. L'utilisation de techniques permettant de quantifier le niveau de transcription de gènes cibles présente donc un intérêt majeur en oncologie. Dans ce domaine, la quantification relative avec la réaction de transcription inverse (*reverse transcription*, RT) suivie d'une PCR quantitative en temps réel (qPCR) est devenue la méthode de choix pour mesurer le niveau d'expression de gènes d'intérêt. La quantification relative consiste à comparer des échantillons inconnus à des échantillons de référence pour mettre en évidence des augmentations ou des diminutions du nombre de transcrits. On peut ainsi estimer l'expression relative d'un gène cible entre différents échantillons biologiques, différentes conditions expérimentales ou différents états physiologiques. En oncologie, les applications sont nombreuses et incluent l'évaluation moléculaire du stade de développement tumoral, ainsi que la mise en évidence de marqueurs pronostiques ou prédictifs de la réponse au traitement.

Sensible, spécifique, et reproductible, la RT-qPCR permet de détecter de faibles changements d'expression des gènes [1]. Cependant, cette technique est loin d'être standardisée, et toute variation dans le protocole utilisé, la qualité de l'ARN, ou l'efficacité de la RT et de la PCR peut facilement introduire des erreurs dans la quantification des transcrits. Dès lors, il est important de normaliser les résultats obtenus

pour corriger la variabilité expérimentale inhérente aux différentes étapes de la réaction. Le choix d'une normalisation appropriée représente l'un des aspects les plus importants de la RT-qPCR à prendre en considération [2]. Malheureusement, si différentes stratégies existent, aucune ne fait l'unanimité. L'objectif de cette revue est de présenter et discuter les différentes options de normalisation disponibles dans le domaine de la cancérologie, afin d'aider à la mise en place d'une stratégie permettant d'obtenir des résultats biologiques pertinents.

## **Normalisation avec l'ARN total**

La normalisation des résultats de RT-qPCR peut se faire par rapport à la quantité d'ARN total utilisée lors de l'étape de transcription inverse. Cette quantité conditionne le rendement de la RT, et doit donc être déterminée précisément. Différentes techniques existent pour mesurer la concentration de la solution d'ARN extrait. La plus commune est de mesurer l'absorbance à 260 nm ( $A_{260}$ ) avec un spectrophotomètre UV. C'est une méthode simple et reproductible. Toutefois, comme une quantité importante d'ARN est requise, les échantillons mesurés sont généralement dilués, ce qui est source d'erreur. De plus, des contaminants absorbants à 260 nm, tels que les protéines, le phénol ou l'ADN génomique, peuvent conduire à des résultats surestimés [3]. A l'heure actuelle se développent de nouveaux systèmes tels que le

NanoDrop<sup>®</sup>. L'avantage majeur de ce spectrophotomètre, dont la sensibilité est estimée à 5 ng/μl, est qu'il ne nécessite qu'1 μl d'échantillon, mis directement en contact avec le système optique. Par ailleurs, l'utilisation d'intercalants fluorescents des acides nucléiques représente une bonne alternative pour quantifier les ARN, et le kit RiboGreen<sup>®</sup> de Molecular Probes repose sur ce principe. C'est une technique plus sensible (seuil de détection : 1 ng/ml) que la mesure de l'absorbance à 260 nm, et qui ne nécessite qu'une faible quantité d'échantillon (1–2 μl). Néanmoins, cette technique ne permet pas de discriminer l'ARN de l'ADN, et des contaminants tels que le phénol peuvent générer des résultats variables [3]. Enfin, le Bioanalyzer 2100 d'Agilent, combiné avec le kit RNA 6000 Nano, permet également de déterminer la concentration de l'ARN. L'analyse n'est pas influencée par des contaminations de phénol ou de protéines, par contre la présence d'ADN génomique nécessite une correction de la mesure de la concentration en ARN [3]. Cependant, la quantification de l'ARN par cette technique n'est pas très précise, et il est préférable de réserver l'appareil à l'analyse qualitative de l'ARN. Les mesures de concentration en ARN variant en fonction des méthodes utilisées, il est indispensable de quantifier tous les échantillons avec la même technique, les systèmes tels que le Nanodrop<sup>®</sup> ou le kit RiboGreen<sup>®</sup> étant préconisés.

Un paramètre important à prendre en compte lorsque l'on normalise par rapport à l'ARN est la qualité de celui-ci. Comme nous l'avons

précédemment rapporté [4], les différences de qualité des échantillons dépendent fortement de l'étape d'extraction, et représentent la source de variations la plus fréquente en RT-qPCR [5]. Aussi est-il primordial d'utiliser la même méthode d'extraction pour tous les échantillons analysés. La qualité de l'ARN est définie à la fois par sa pureté (absence de contaminants) et son intégrité (ARN non dégradé). Sa pureté est déterminée par la mesure de l'absorbance à 230 (contaminants organiques) et 280 nm (spécifique des protéines), l'ARN étant considéré pur pour des ratios  $A_{260}/A_{230}$  et  $A_{260}/A_{280} > 1.8$ . En ce qui concerne l'intégrité, la méthode traditionnelle consistait à visualiser les bandes d'ARN ribosomiques 28S et 18S sur un gel d'électrophorèse [4]. En effet, il est difficile d'analyser directement les ARNm, ceux-ci ne représentant qu'1% à 3% des ARN totaux. On doit donc considérer que la dégradation des ARN ribosomiques, majoritaires, reflète la dégradation des ARNm. Ainsi, le ratio 18S/28S évalue l'intégrité de l'ARN, un ratio proche de 2 étant considéré comme indicateur d'un ARN peu ou non dégradé. Cependant, cette méthode nécessite une importante quantité d'ARN (0.5–2 mg), et n'est pas assez sensible pour détecter de légères dégradations [6]. Récemment, l'analyse qualitative de l'ARN a été grandement simplifiée et améliorée avec le développement de systèmes d'électrophorèse capillaire sur puce, tels que le Bioanalyzer 2100 d'Agilent. Cette technique permet de qualifier les ARN dont la concentration est comprise entre 5 et 500 ng/ $\mu$ l. Pour chaque échantillon, le logiciel détermine le ratio 28S/18S, et attribue un RIN (RNA Integrity Number) qui tient compte de l'intégralité de

l'électrophorogramme [7]. La valeur du RIN s'étend de 1 à 10, 1 correspondant à un ARN totalement dégradé, et 10 à un ARN de très bonne qualité. En plus d'être rapide et de permettre un haut débit, cette technique ne nécessite qu'1 µl d'échantillon. Elle représente la méthode la plus simple et objective pour l'analyse qualitative de l'ARN ; son utilisation est donc recommandée [5].

Un dernier point important à prendre en compte est la présence éventuelle d'inhibiteurs. D'un point de vue mathématique, la réaction de transcription inverse peut être représentée par l'équation suivante :  $[ADNc] = [ARN] \times \text{Efficacité}$ . Il n'y a pas d'amplification, et l'unique variable est l'efficacité, qui représente le pourcentage d'ARNm convertis en ADNc [8]. Cette efficacité peut varier de 5 à 90% [8], car la réaction est extrêmement sensible à la présence d'inhibiteurs, qui peuvent être des réactifs utilisés dans l'étape d'extraction (sels, alcools, phénol), ou des composants biologiques copurifiés (urée, hème, héparine, immunoglobuline G) [8,9]. Ces composés peuvent aussi inhiber la réaction de PCR. Dès lors, 2 réactions avec une quantité égale d'ARN, mais dont les efficacités de RT et/ou de PCR sont différentes, donneront des résultats qui ne pourront pas être comparés. Différentes méthodes existent pour évaluer la présence d'inhibiteurs dans les échantillons biologiques. Tout d'abord, il est possible de comparer l'efficacité de PCR pour différentes dilutions (1/20 et 1/80 par exemple) d'un échantillon. Une alternative est d'ajouter une quantité définie d'un amplicon synthétique simple brin à l'ADNc des échantillons,

et de comparer son amplification par rapport à un contrôle sans ADNc [9,10]. Néanmoins, ces deux méthodes se limitent à vérifier l'absence d'inhibiteurs de PCR, et ne permettent pas d'évaluer l'efficacité de l'étape de transcription inverse. Une solution intéressante pour contrôler les deux réactions enzymatiques (RT et PCR) consiste à ajouter à l'ARN extrait un ARN exogène, dont on comparera l'amplification entre les différents échantillons [11]. Cette séquence contrôle ne doit présenter aucune similitude avec l'ARN cible ; on utilisera par exemple un ARNm spécifique d'une plante lorsqu'on étudie l'expression génique chez l'homme.

Certaines études ont montré que la normalisation par rapport à l'ARN total était l'approche la moins incertaine [1,12], et qu'elle devrait être considérée comme appropriée lorsque des biopsies constituent le matériel biologique de départ [10]. Cependant, cette approche présume que l'ARNm représente une proportion constante de l'ARN total dans toutes les cellules. Or, l'activité transcriptionnelle peut varier, et on peut dès lors avoir des déséquilibres dans la proportion des différents types d'ARN (ARNr, ARNt et ARNm) entre les individus, ou entre tissu normal et tumoral [13].

## **Normalisation avec le nombre de cellules**

Normaliser le niveau d'expression des ARNm par rapport au nombre de cellules des échantillons étudiés est une stratégie attractive. Mais si cela est envisageable pour les prélèvements sanguins et les cultures cellulaires, cette approche n'est pas applicable aux tissus et aux tumeurs solides [1], pour lesquels la quantité de cellules ne peut être déterminée de façon précise. Pour contourner ce problème, il a été suggéré de normaliser les données de RT-qPCR entre les échantillons en utilisant la quantité d'ADN génomique comme indicateur reflétant le nombre de cellules de chaque échantillon [14]. Cependant, ces deux approches ne tiennent pas compte de la dégradation de l'ARN, ni de l'efficacité de la RT et de la PCR. Par ailleurs, dans le cas de la normalisation avec l'ADN génomique, le fait de travailler avec des cellules tumorales peut présenter des problèmes supplémentaires car elles présentent généralement des anomalies du caryotype. Dès lors, le rapport entre la quantité d'ADN et le nombre de cellules devient variable.

## **Normalisation avec des gènes de référence**

La normalisation par rapport à un gène de référence est la plus couramment utilisée. Les gènes de référence, ou gènes de ménage, sont un groupe de gènes qui codent pour des protéines dont la fonction

est essentielle à la viabilité de la cellule. Il est généralement présumé que ces gènes ont une expression ubiquitaire, stable entre les différents tissus, entre les individus, et quelles que soient les conditions expérimentales. L'utilisation de ce contrôle endogène permet en théorie de contrôler toutes les étapes du protocole expérimental, son expression reflétant non seulement la quantité et la qualité de l'ARN utilisé, mais aussi les efficacités de la RT et de la PCR.

### **Les modèles existants**

La concentration initiale d'une cible est généralement dérivée de son  $C_T$  (*Cycle threshold*), qui correspond au nombre de cycles d'amplification où la courbe d'amplification croise la ligne de seuil (*figure 1*). Cette ligne est placée au niveau de la phase exponentielle, de façon à se distinguer clairement du bruit de fond. Pour chaque échantillon analysé, les  $C_T$  obtenus pour les gènes d'intérêt et de référence doivent être convertis en ratio d'expression normalisé. Pour cela, différentes options sont disponibles, qu'elles soient intégrées dans les logiciels fournis avec les différents appareils de qPCR, ou décrites dans la littérature [15,16,18,19].

La méthode relative des courbes standard [15] nécessite la construction, pour le gène cible et le gène de référence, d'une gamme réalisée à partir d'une série de dilutions d'un échantillon de référence. Ces gammes permettent d'obtenir des courbes standard, obtenues en exprimant les  $C_T$  en fonction du Log de la concentration initiale d'ADNc.

Les valeurs de concentration pour chaque point de la gamme peuvent être fixées de façon arbitraire en respectant les facteurs de dilution. Dès lors, la quantité relative d'une cible est déterminée grâce à son  $C_T$ , par interpolation avec la courbe standard. L'expression normalisée d'un gène d'intérêt est donc déterminée par la formule suivante :

$$R = \frac{\text{Quantité relative du gène d'intérêt}}{\text{Quantité relative du gène de référence}}$$

Par ailleurs, un calibrateur est généralement utilisé. Il s'agit d'un échantillon utilisé comme base pour comparer les résultats. Le ratio normalisé de chaque échantillon est divisé par le ratio normalisé du calibrateur. Ainsi, le calibrateur devient la référence 1x, et tous les autres échantillons sont exprimés sous forme de ratio par rapport au calibrateur.

La méthode du  $\Delta\Delta C_T$  [16] utilise une formule arithmétique pour calculer le ratio d'expression d'un gène cible entre 2 échantillons, en normalisant avec un gène de référence. Tout d'abord, les différences  $\Delta C_T$  entre les valeurs de  $C_T$  du gène cible et du gène de référence sont déterminées pour l'échantillon à analyser et le contrôle.

$$\Delta C_{T(\text{éch})} = C_T(\text{cible}_{\text{éch}}) - C_T(\text{référence}_{\text{éch}})$$

$$\Delta C_{T(\text{cont})} = C_T(\text{cible}_{\text{cont}}) - C_T(\text{référence}_{\text{cont}})$$

Ensuite, le  $\Delta\Delta C_T$  entre le contrôle et l'échantillon est calculé:

$$\Delta\Delta C_T = \Delta C_{T(\text{cont})} - \Delta C_{T(\text{éch})}$$

Enfin, le ratio d'expression normalisé d'un gène cible est déterminé par la formule  $2^{-\Delta\Delta C_T}$ .

Contrairement à la méthode relative des courbes standard, où l'efficacité d'amplification (E) des gènes cibles et de référence est directement prise en compte lors de la construction des gammes, la méthode du  $\Delta\Delta C_T$  présume que les efficacités des 2 gènes sont égales à 100% ( $E = 2$ , à chaque cycle de la phase exponentielle, la concentration en produits de PCR est doublée).

Cependant, une différence d'efficacité de PCR de 3% ( $\Delta E = 0.03$ ) entre les 2 gènes génère une erreur de 47% pour le ratio d'expression si  $E_{\text{cible}} < E_{\text{ref}}$  et 209% si  $E_{\text{cible}} > E_{\text{ref}}$  après 25 cycles [17]. De plus, cette erreur augmente exponentiellement avec de plus grandes variations d'efficacité et un plus grand nombre de cycles.

De nouveaux modèles ont été développés en prenant en compte l'efficacité de PCR du gène cible et du gène de référence. Le plus répandu est le modèle de Pfaffl [18], où le ratio d'expression relative (R) d'un gène cible entre un échantillon et un contrôle est déterminé par la formule suivante :

$$R = \frac{\Delta C_{T \text{ cible}} (\text{contrôle} - \text{échantillon})}{\Delta C_{T \text{ ref}} (\text{contrôle} - \text{échantillon})} \cdot \frac{(E_{\text{cible}})^{\Delta C_{T \text{ cible}} (\text{contrôle} - \text{échantillon})}}{(E_{\text{ref}})^{\Delta C_{T \text{ ref}} (\text{contrôle} - \text{échantillon})}}$$

Dans ce modèle de Pfaffl, l'efficacité de PCR pour un gène donné est calculée à partir de la construction d'une courbe de calibration en utilisant la formule suivante :  $E = 10^{[-1/\text{pente}]}$ . Cette méthode donne une bonne estimation de l'efficacité, même s'il est possible qu'elle soit

surestimée [17]. Toutefois, cette approche présume que les efficacités d'amplification entre les échantillons dilués sont identiques, créant ainsi une relation linéaire entre le  $C_T$  et la quantité d'ADNc au départ.

Dès lors, certains auteurs comme Liu et Saint [19] ont développé des modèles de normalisation qui tiennent compte de l'efficacité propre à chaque échantillon, celle-ci étant déterminée par la cinétique de la courbe d'amplification. Toutefois, avec ce genre d'approche, la moindre erreur dans la mesure de l'efficacité est amplifiée et répercutée de façon exponentielle sur le ratio d'expression calculé [20].

Les différents modèles de normalisation avec les gènes de référence présentent donc tous des avantages et des inconvénients. A l'heure actuelle, il n'existe pas de méthode consacrée pour le traitement des résultats de RT-qPCR.

### **Quel gène de référence choisir ?**

La fiabilité des données normalisées dépend de la robustesse du gène contrôle. Idéalement, ce gène doit avoir une expression constitutive, non régulée, stable entre les différents échantillons. Généralement, des gènes de la maintenance cellulaire sont utilisés pour la normalisation (*tableau 1*). Ces gènes régulent des fonctions cellulaires de base, et codent par exemple pour des composants du cytosquelette ( $\beta$ -actine, ACTB), le complexe majeur d'histocompatibilité ( $\beta$ -2-microglobuline, B2M), la voie de la glycolyse (glyceraldehyde-3-phosphate dehydrogenase, GAPDH, phosphoglycerokinase 1, PGK1), le recyclage métabolique des nucléotides (hypoxanthine ribosyltransferase, HPRT),

ou la synthèse des sous-unités ribosomiques (ARNr, 18S ou 28S). Dans de nombreuses études, l'expression de ces gènes est présumée invariable entre les cellules des différents échantillons, et ils sont utilisés pour la normalisation. Cependant, la plupart des gènes de référence communément utilisés sont régulés, et ont une expression variable [1,5]. En effet, ces gènes de ménage ne sont pas seulement impliqués dans le métabolisme cellulaire basal, mais ils participent aussi à d'autres fonctions. GAPDH par exemple est impliqué non seulement dans la glycolyse, mais également dans de nombreux processus cellulaires tels que la réparation de l'ADN, la maintenance des télomères, ou l'apoptose [21]. Les concentrations de GAPDH peuvent varier significativement entre différents individus [22], entre différents tissus [23], entre tissu sain et tumoral [10,24,25], ou sous l'influence d'un traitement expérimental [26]. Ainsi, de nombreuses études montrent que les gènes de ménage couramment utilisés sont régulés, et qu'aucun n'est exprimé de façon constitutive dans tous les types cellulaires et sous toutes les conditions expérimentales [2]. Certains auteurs considèrent que ces problèmes de variabilité peuvent être contournés en normalisant avec les ARN ribosomiques (ARNr), qui semblent moins affectés par les conditions expérimentales [26]. Ainsi, l'ARNr 18S est souvent utilisé comme référence, et diverses études ont montré qu'il était exprimé à un niveau relativement constant dans diverses localisations cancéreuses [27-30]. Cette stabilité pourrait résulter du fait que l'ARN 18S n'est pas impliqué dans le métabolisme cellulaire [31]. Toutefois, d'autres études ont montré que l'ARN 18S

était régulé [13] ou considèrent ce gène comme un contrôle inapproprié car il est moins affecté par la dégradation que les ARNm [30]. De ce point de vue, l'ARN 28S est considéré plus représentatif de l'intégrité de l'ARN. Cependant, Steinau *et al.* [11] ont montré que l'expression de l'ARN 28S pouvait varier. Par ailleurs, d'autres limitations à l'utilisation des ARNr peuvent être rapportées. En effet, les ARNr sont transcrits avec une polymérase différente des ARNm, donc les variations de l'activité polymérase pourraient ne pas affecter l'expression des 2 types d'ARN de façon similaire [32]. De plus, ces ARNr sont très abondants (en général les  $C_T$  obtenus sont  $< 10$  [29]), il est dès lors nécessaire de diluer les échantillons, ce qui augmente les risques d'erreur [33].

Il apparaît donc que le choix d'un contrôle approprié est une question délicate. Malgré cela, dans de nombreuses publications, les gènes de ménage continuent à être utilisés pour normaliser les données de RT-qPCR sans que leur validité ait été démontrée [2,34]. Pour contrer ces problèmes, de nouvelles méthodes de normalisation, proposant l'utilisation de multiples gènes de référence, ont émergé [35-37]. Dès lors, des outils tels que geNorm [36] ou BestKeeper [35] ont été mis à la disposition de la communauté scientifique. Ces programmes permettent d'identifier, parmi un panel de gènes testés, une combinaison de gènes de référence appropriée à une expérience donnée. Cependant, peut-on considérer constante une moyenne de plusieurs gènes de référence ? Cette approche sous-estime l'énorme variabilité observée entre les individus lorsqu'on travaille sur des tissus humains [2]. Une stratégie plus intéressante est d'identifier un gène de référence adaptée à un

type cellulaire, une pathologie précise, ou des conditions expérimentales données.

Nous avons recensé dans le tableau 2 les travaux qui ont étudié la stabilité de gènes de ménage par RT-qPCR dans diverses localisations cancéreuses. Bien que certains [29,38] aient recherché un gène de référence universel, exprimé à un niveau constant quelque soit le tissu, la plupart ont essayé de valider un ou plusieurs contrôles spécifiques à une localisation précise. Néanmoins, dans ce genre d'études, si l'on veut évaluer la stabilité des gènes de référence de manière pertinente, il est important de limiter les différences introduites par des variations de la quantité et de la qualité de l'ARN utilisé, et par les variations d'efficacité de la RT et de la PCR. Toutes les précautions recommandées pour la normalisation avec la quantité d'ARN total doivent donc être appliquées. On remarquera cependant que peu d'auteurs ont apporté une attention particulière à ces paramètres importants que sont la quantification de l'ARN, sa qualification, et la vérification de l'absence d'inhibiteurs. Concernant la mesure de la concentration de l'ARN par exemple, seules 2 études ont utilisé des techniques précises tels que le kit Ribogreen® [39] ou le système Nanodrop® [25]. Dans certaines études, la méthode utilisée n'a pas été indiquée [28,29,40], ou bien était connue pour être assez imprécise [24]. A l'extrême, Steinau *et al.* [11] ont utilisé des quantités d'ARN différentes lors de la transcription inverse. C'est le volume d'ARN qui était identique, mais les concentrations étaient différentes. Et bien que le nombre de cellules pour chaque échantillon ait été évalué, cela n'a

pas servi à la normalisation des résultats pour les gènes de référence, mais seulement à la corrélation entre ces résultats et la « quantité de cellules ».

Par ailleurs, seul un nombre restreint d'études a contrôlé la présence d'inhibiteurs. Néanmoins, le système mis en place se limitait souvent au contrôle de l'inhibition de la PCR [10,29,34]. Seuls Steinau *et al.* [11], en utilisant un ARN exogène, vérifiaient l'inhibition de la RT et de la PCR.

Enfin, si les études recensées dans le tableau 2 identifient les gènes les moins variables, cela ne veut pas forcément dire qu'ils soient stables. Il peut en effet y avoir 2 cycles d'écart entre les  $C_T$  obtenus pour les différents échantillons avec le gène le moins variable [41]. Dès lors, les divergences de protocole expérimental et de méthode d'analyse peuvent mener à des conclusions différentes dans le choix d'un gène de référence approprié. En effet, sur 21 gènes analysés, Rubie *et al.* [33] ont identifié un panel de 3 gènes montrant des variations minimales entre colon normal et tumoral (PMM1, ACTB et PSMB6), et ces gènes ont été recommandés pour la normalisation dans cette localisation. A l'inverse, HPRT était le gène qui montrait la plus grande instabilité, et son utilisation était donc déconseillée. Pourtant, une autre étude menée par de Kok *et al.* [34] a conclu que l'utilisation de HPRT était appropriée pour la normalisation dans différents tissus, y compris le colon. Cependant, l'approche utilisée dans cette étude est critiquable. En effet, les auteurs ont émis l'hypothèse que l'expression moyenne de 11 gènes de ménages était stable, sans vraiment l'avoir vérifié. Parmi ce

panel, ils ont alors montré que HPRT était le gène qui, pour tous les tissus épithéliaux analysés, reflétait au mieux l'expression moyenne des 11 autres.

Ce genre de discordance souligne l'importance pour les expérimentateurs de valider leurs gènes de référence dans leur propre modèle. Pour cela, il est important de mettre en place un maximum de contrôles et d'optimiser chaque étape du protocole. Notamment, il est important de quantifier l'ARN de façon précise, de s'assurer de son intégrité avec un système d'électrophorèse capillaire sur puce, et de vérifier l'efficacité de la RT et de la PCR à l'aide d'un ARN exogène. Enfin, il est intéressant de vérifier la variabilité intrinsèque des résultats. C'est ce qu'ont fait par exemple Lossos *et al.* [30]. L'expression de plusieurs gènes a été évaluée pour 8 aliquots d'une même biopsie, traités comme des échantillons distincts, et tous les gènes ont montré une très faible variance inter-essais.

## Conclusion

Bien que fiable et robuste, la technique de qPCR peut générer des résultats variables lorsqu'elle est appliquée à la quantification des transcrits. La labilité de l'ARN, la difficulté de le quantifier précisément, et la présence possible d'inhibiteurs sont autant de facteurs qui peuvent engendrer des variations susceptibles de fausser les résultats. Il est donc important de définir une stratégie de normalisation des données

obtenues, en prenant en compte les différentes considérations abordées dans cette revue. Différentes approches sont envisageables, et celles-ci étant non exclusives les unes des autres, il paraît intéressant de les utiliser de façon complémentaire.

## Références

1. Bustin SA. Absolute quantification of mRNA using real-time reverse transcription polymerase chain reaction assays. *J Mol Endocrinol* 2000 ; 25 : 169-93.
2. Bustin SA, Benes V, Nolan T, Pfaffl MW. Quantitative real-time RT-PCR - a perspective. *J Mol Endocrinol* 2005 ; 34 : 597-601.
3. Lightfoot S. Quantitation comparison of total RNA using the agilent 2100 Bioanalyzer, Ribogreen analysis and UV spectrometry. Agilent Technologies - Application Note 5988-7650EN 2006.
4. Bastard J P, Chambert S, Ceppa F, Coude M, Crapez E, Loric S, Muzeau F, Spyrtos F, Poirier K, Copois V, Tse C, Bienvenu T. RNA isolation and purification methods. *Ann Biol Clin* 2002 ; 60 : 513-23.
5. Bustin SA, Nolan T. Pitfalls of Quantitative real-time reverse-transcription polymerase chain reaction. *J Biomol Tech* 2004 ; 15 : 155-66.
6. Imbeaud S, Graudens E, Boulanger V, Barlet X, Zaborski P, Eveno E, Mueller O, Schroeder A, Auffray C. Towards standardization of RNA quality assessment using user-independent classifiers of microcapillary electrophoresis traces. *Nucleic Acids Res* 2005 ; 33 : e56.
7. Schroeder A, Mueller O, Stocker S, Salowsky R, Leiber M, Gassmann M, Lightfoot S, Menzel W, Granzow M, Ragg T. The RIN: an

RNA integrity number for assigning integrity values to RNA measurements. *BMC Mol Biol* 2006 ; 7 : 3.

8. Freeman WM, Walker SJ, Vrana KE. Quantitative RT-PCR: pitfalls and potential. *Biotechniques* 1999 ; 26 : 112-5.

9. Nolan T, Hands RE, Ogunkolade W, Bustin SA. SPUD: a quantitative PCR assay for the detection of inhibitors in nucleic acid preparations. *Anal Biochem* 2006 ; 351 : 308-10.

10. Tricarico C, Pinzani P, Bianchi S, Paglierani M, Distante V, Pazzagli M, Bustin SA, Orlando C. quantitative real-time reverse transcription polymerase chain reaction: normalization to rRNA or single housekeeping genes is inappropriate for human tissue biopsies. *Anal Biochem* 2002 ; 309 : 293-300.

11. Steinau M, Rajeevan MS, Unger ER. DNA and RNA references for qRT-PCR assays in exfoliated cervical cells. *J Mol Diagn* 2006 ; 8 : 113-8.

12. Bustin SA. Quantification of mRNA using real-time reverse transcription PCR (RT-PCR): trends and problems. *J Mol Endocrinol* 2002 ; 29 : 23-39.

13. Solanas M, Moral R, Escrich E. Unsuitability of using ribosomal RNA as loading control for northern blot analyses related to the imbalance between messenger and ribosomal RNA content in rat mammary tumors. *Anal Biochem* 2001 ; 288 : 99-102.

14. Kanno J, Aisaki K, Igarashi K, Nakatsu N, Ono A, Kodama Y, Nagao T. "Per cell" normalization method for mRNA measurement by quantitative PCR and microarrays. *BMC Genomics* 2006 ; 7 : 64-77.
15. ABI. Relative quantitation of gene expression. User bulletin No. 2. ABI prism 7700 Sequence Detection System. PE Applied Biosystems.
16. 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 : 402-8.
17. Pfaffl MW. Quantification strategies in real time PCR. *A-Z of quantitative PCR*. La Jolla, CA : International UniversityLine, 2004 ; 87-120.
18. Pfaffl MW. A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res* 2001 ; 29 : e45.
19. Liu W, Saint DA. A new quantitative method of real time reverse transcription polymerase chain reaction assay based on simulation of polymerase chain reaction kinetics. *Anal Biochem* 2002 ; 302 : 52-9.
20. Peirson SN, Butler JN, Foster RG. Experimental validation of novel and conventional approaches to quantitative real-time PCR data analysis. *Nucleic Acids Res* 2003 ; 31 : e73.
21. Sirover MA. New nuclear functions of the glycolytic protein, glyceraldehyde-3-phosphate dehydrogenase, in mammalian cells. *J Cell Biochem* 2005 ; 95 : 45-52.

22. Bustin SA, Gyselman VG, Williams NS, Dorudi S. Detection of cytokeratins 19/20 and guanylyl cyclase C in peripheral blood of colorectal cancer patients. *Br J Cancer* 1999 ; 79 : 1813-20.
23. Barber RD, Harmer DW, Coleman RA, Clark BJ. GAPDH as a housekeeping gene: analysis of GAPDH mRNA expression in a panel of 72 human tissues. *Physiol Genomics* 2005 ; 21 : 389-95.
24. Silvia S, Francesca C, Marco LI, Silvia N, Giorgio VS, Raffaele C A. Selection of suitable reference genes for accurate normalization of gene expression profile studies in non-small cell lung cancer. *BMC Cancer* 2006 ; 6 : 200-9.
25. Ohl F, Jung M, Xu C, Stephan C, Rabien A, Burkhardt M, Nitsche A, Kristiansen G, Loening SA, Radonic A, Jung K. Gene expression studies in prostate cancer tissue: which reference gene should be selected for normalization ? *J Mol Med* 2005 ; 83 : 1014-24.
26. Schmittgen TD, Zakrajsek BA. Effect of experimental treatment on housekeeping gene expression: validation by real-time, quantitative RT-PCR. *J Biochem Biophys Methods*. 2000 ; 46 : 69-81.
27. Blanquicett C, Johnson MR, Heslin M, Diasio RB. Housekeeping gene variability in normal and carcinomatous colorectal and liver tissues: applications in pharmacogenomic gene expression studies. *Anal Biochem* 2002 ; 303 : 209-14.
28. Goidin D, Mamessier A, Staquet MJ, Schmitt D, Berthier-Vergnes O. Ribosomal 18S RNA prevails over glyceraldehyde-3-phosphate

dehydrogenase and beta-actin genes as internal standard for quantitative comparison of mRNA levels in invasive and noninvasive human melanoma cell subpopulations. *Anal Biochem* 2001 ; 295 : 17-21.

29. Aerts JL, Gonzales MI, Topalian SL. Selection of appropriate control genes to assess expression of tumor antigens using real-time RT-PCR. *Biotechniques* 2004 ; 36 : 84-91.

30. Lossos IS, Czerwinski DK, Wechser MA, Levy R. Optimization of quantitative real-time RT-PCR parameters for the study of lymphoid malignancies. *Leukemia* 2003 ; 17 : 789-95.

31. Finnegan MC, Goepel JR, Hancock BW, Goyns MH. Investigation of the expression of housekeeping genes in non-Hodgkin's lymphoma. *Leuk Lymphoma* 1993 ; 10 : 387-93.

32. Radonic A, Thulke S, Mackay IM, Landt O, Siegert W, Nitsche A. Guideline to reference gene selection for quantitative real-Time PCR. *Biochem Biophys Res Commun* 2004 ; 313 : 856-62.

33. Rubie C, Kempf K, Hans J, Su T, Tilton B, Georg T, Brittner B, Ludwig B, Schilling M. Housekeeping gene variability in normal and cancerous colorectal, pancreatic, esophageal, gastric and hepatic tissues. *Mol Cell Probes* 2005 ; 19 : 101-9.

34. de Kok JB, Roelofs RW, Giesendorf BA, Giesendorf BA, Pennings JL, Waas ET, Feuth T, Swinkels DW, Span PN. Normalization of gene

expression measurements in tumor tissues: comparison of 13 endogenous control genes. *Lab Invest* 2005 ; 85 : 154-9.

35. 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 : 509-15.

36. 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 : research0034.1-0034.11

37. 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 : 5245-50.

38. Gerard CJ, Andrejka LM, Macina RA. Mitochondrial ATP synthase 6 as an endogenous control in the quantitative RT-PCR analysis of clinical cancer samples. *Mol Diagn* 2000 ; 5 : 39-46.

39. Dydensborg AB, Herring E, Auclair J, Tremblay E, Beaulieu JF. Normalizing genes for quantitative RT-PCR in differentiating human intestinal epithelial cells and adenocarcinomas of the colon. *Am J Physiol Gastrointest Liver Physiol* 2006 ; 209 : 1067-74

40. Lupberger J, Kreuzer KA, Baskaynak G, Peters UR, le Coutre P, Schmidt CA. Quantitative analysis of beta-actin, beta-2-microglobulin

and porphobilinogen deaminase mRNA and their comparison as control transcripts for RT-PCR. *Mol Cell Probes* 2002 ; 16 : 25-30.

41. Liu DW, Chen ST, Liu HP. Choice of endogenous control for gene expression in nonsmall cell lung cancer. *Eur Respir J* 2005 ; 26 : 1002-8.

42. Fischer M, Skowron M, Berthold F. Reliable transcript quantification by real-time reverse transcriptase-polymerase chain reaction in primary neuroblastoma using normalization to averaged expression levels of the control genes HPRT1 and SDHA. *J Mol Diagn* 2005 ; 7 : 89-96.

43. Neuvians TP, Gashaw I, Sauer CG, von Ostau C, Kliesch S, Bergmann M, Hacker A, Grobholz R. Standardization strategy for quantitative PCR in human seminoma and normal testis. *J Biotechnol* 2005 ; 117 : 163-71.