

## UNA ASIGNATURA PENDIENTE: LA NORMALIZACIÓN DE LA RT-qPCR

### A PENDING SUBJECT: NORMALIZATION OF RT-qPCR

por RAFAEL A. CAÑAS, M<sup>a</sup> BELÉN PASCUAL Y FERNANDO DE LA TORRE

INVESTIGADORES DEL DEPARTAMENTO DE BIOLOGÍA MOLECULAR Y BIOQUÍMICA. FACULTAD DE CIENCIAS. UNIVERSIDAD DE MÁLAGA.

rcanas@uma.es, bpascual@uma.es, fdelatorre@uma.es

*Palabras clave: expresión génica, RT-qPCR, normas MIQE, genes normalizadores*  
*Keywords: gene expression, RT-qPCR, MIQE guidelines, normalization genes*

Enviado: 9 de noviembre de 2016  
Aceptado: 12 de diciembre de 2016

En los últimos años, la biología molecular ha pasado por una serie de revoluciones técnicas. La RT-qPCR es una técnica con un papel principal en una de estas revoluciones. Desde su aparición, los análisis de expresión génica se han vuelto cuantitativos, más baratos y más rápidos, por lo que son más populares en la comunidad científica, difundiendo su uso por todos los laboratorios. Sin embargo, la técnica no es tan sencilla como la mayoría de los investigadores piensan. La incorrecta implementación de la RT-qPCR conduce a resultados inconsistentes e irrepetibles. Sin embargo, incluso en este momento, hay investigadores haciendo esfuerzos para reorientar esta situación estableciendo una serie de directrices como en el caso de las normas MIQE. Uno de los temas más problemáticos es la estimación de la expresión génica mediante la normalización de los resultados crudos de RT-qPCR. Por lo general, casi todos los investigadores consideran que la normalización se debe realizar utilizando genes de referencia con expresión estable en cada condición o tejido, los denominados genes «constitutivos». Sin embargo, este concepto ha sido superado por la experiencia que muestra que la selección de los genes de normalización debe lograrse a través de procedimientos experimentales siguiendo un «código de buena conducta» como las directrices MIQE.

*In recent years, the molecular biology has gone through a series of technical revolutions. The RT-qPCR is a technique with a leading role in one of these revolutions. Since its emergence, the gene expression analyses have become quantitative, cheaper and faster, thus more popular in the researchers' community spreading its use across all the laboratories. However, the technique is not as easy as the most of researchers think. The incorrect implementation of the RT-qPCR leads to inconsistent and unrepeatable results. Nevertheless, even at this point, there are researchers making efforts to redirect this situation establishing a series of guidelines as in the case of MIQE standards. One of the most problematic issues is the estimation of the gene expression normalizing the raw results from a RT-qPCR. Usually, almost every researcher consider that the normalization must be performed using reference genes with stable expression in every condition or tissue, so-called "housekeeping" genes. However, this concept has been overcome by the experience showing that the selection of the normalization genes must be achieved through experimental procedures following a "code of good conduct" such as the MIQE guidelines.*

La reacción de la polimerasa en cadena (PCR, siglas en inglés) es una de las tecnologías que han cambiado el mundo en los últimos 30 años. La PCR confiere a la humanidad la capacidad de realizar la replicación y amplificación *in vitro* del ácido desoxirribonucleico (DNA, siglas en inglés), convirtiéndose la manipulación y análisis de la molécula de la herencia en algo rutinario. Esto ha permitido un desarrollo extraordinario de la Biología Molecular, modificando la forma de trabajar en numerosas disciplinas biológicas y clínicas. Un ejemplo muy atractivo son los actuales métodos de secuenciación masiva del DNA que nunca se hubiesen podido desarrollar sin la PCR.

La principal fortaleza de la PCR está en su versatilidad a partir de la cual se han podido desarrollar numerosas variantes de la técnica primigenia con ob-

jetivos muy dispares (PCR anidada, PCR cuantitativa, PCR *in situ*, RT-PCR, RACE-PCR... ). Quizá, la variante más extendida y de mayor impacto en la actualidad es la reversotranscripción seguida de una PCR cuantitativa (RT-qPCR, siglas en inglés). Esta técnica permite conocer el nivel de expresión de un gen (transcripción) a partir de la estimación de la cantidad de ácido ribonucleico mensajero (mRNA, siglas en inglés) de una muestra biológica. Por lo tanto, esta técnica ha sustituido en la rutina del laboratorio de Biología Molecular a la clásica técnica de *northern blot*, basada en la hibridación de moléculas de ácidos nucleicos y con una larga serie de inconvenientes (largo y tedioso desarrollo, menor fiabilidad de los resultados, frecuente uso de revelados radioactivos, resultados no cuantitativos...). El

principio de la RT-qPCR es muy simple. A partir de una muestra de RNA se sintetiza un DNA copia (cDNA), el cual será utilizado como molde para desarrollar una PCR a tiempo real (real-time PCR, siglas en inglés) empleando oligonucleótidos con secuencias específicas del gen cuya expresión se quiere determinar. Una de las principales ventajas de la PCR a tiempo real es que la cantidad de DNA producido puede ser monitorizada en todo momento. Para ello, se mide la fluorescencia emitida por sondas específicas de DNA unidas a fluorocromos o por compuestos químicos que al intercalarse entre las bases del DNA emiten fluorescencia. A partir de esta información se obtienen los datos que permiten, previo procesamiento matemático, estimar la cantidad de mRNA en una muestra problema. Lo que permite que esta información pueda ser realmente cuantitativa (absoluta) o semicuantitativa (relativa). Normalmente, en la práctica de laboratorio, el método más usado es el semicuantitativo, puesto que el cuantitativo requiere de curvas de calibración a partir de patrones de DNA que son muy dependientes de las condiciones en las que se lleva a cabo la técnica y de su análisis posterior. Para el análisis semicuantitativo se requiere uno o más genes de referencia cuya expresión sea estable en todas las muestras analizadas en un mismo experimento. Son los datos de estos genes los que se usan para normalizar matemáticamente la expresión de los genes que se quieren analizar.

El proceso de elaboración y análisis de datos mediante esta técnica puede ser realmente rápido permitiendo conocer la expresión de una decena de genes en cientos de muestras diferentes en tan solo uno o dos días. Esto provoca a veces una especie de ofuscación en el investigador ante la perspectiva de obtener una buena cantidad de resultados en poco tiempo. La materialización de este estado lleva al descuido, y a veces a la negligencia, de ciertos aspectos en la preparación y aplicación de la RT-qPCR. Esto ha conllevado la publicación de resultados falsos, no fiables o que nunca podrán ser repetidos experimentalmente<sup>[1]</sup>. Tal ha sido y es la dimensión del problema que varios grupos de investigadores han propuesto diferentes métodos de estandarización (a nivel metodológico) de la RT-qPCR para una correcta aplicación de la técnica que permita la publicación de datos reproducibles y confiables. Entre estas propuestas la más aceptada se ha llamado «Información Mínima para la Publicación de Experimentos de PCR a Tiempo-Real Cuantitativa» (MIQE, siglas en inglés)<sup>[2]</sup>.

Lo extendido del uso de la RT-qPCR en su modalidad semicuantitativa provoca que la determinación y uso de genes de referencia sea uno de los prin-

cipales problemas en la aplicación de esta técnica. Conceptualmente el gen de referencia debe tener una expresión constitutiva en las muestras que se quieren analizar, por lo que los cambios en las medidas de la RT-qPCR han de deberse a variaciones experimentales, como la calidad del RNA de partida o la reacción de reversotranscripción y nunca a los tratamientos aplicados sobre las muestras. En principio, lo mejor sería tener genes de referencia «universales», cuya expresión fuese la misma en cualquier circunstancia. Esto permitiría ahorrar trabajo y comparar, siempre de la misma forma, los resultados para todas las muestras y genes problema. Por ello se acudió al concepto de genes «*housekeeping*», aquellos que mantienen una expresión constitutiva por estar encargados de las funciones celulares básicas. Así los investigadores se lanzaron a usar genes considerados como «*housekeeping*» para la normalización de las RT-qPCR, en muchos casos sin ningún tipo de comprobación experimental. Con el tiempo incluso se han llegado a hacer listas de genes de normalización/«*housekeeping*» clásicos (*factor de elongación 1 $\alpha$* , *tubulina*, *actina*, *gliceraldehído 3 fosfato deshidrogenasa*... ). El problema ha llegado cuando se ha comprobado que el concepto de gen «*housekeeping*» no está tan claro, observándose que muchos de los genes catalogados en esta categoría no presentan una expresión constitutiva en todas las condiciones experimentales. Uno de los casos más famosos es el de la *gliceraldehído 3 fosfato deshidrogenasa*, que después de ser usado como gen de referencia durante muchos años, se ha comprobado que posee expresión diferencial en diferentes tejidos o muestras. Por ejemplo, en los diferentes tejidos humanos<sup>[3]</sup> y en respuesta a la infección por *Mycobacterium tuberculosis*<sup>[4]</sup>. Esto ha provocado que se empiece a considerar que un gen puede ser usado como referencia en una RT-qPCR siempre y cuando se haya confirmado previamente que su expresión es estable en las muestras a analizar, independientemente de cual sea su expresión en otros tejidos o condiciones experimentales. Es ahí donde entra en juego la realización de experimentos previos para validar el uso de un gen como referencia. Adicionalmente se han diseñado nuevas herramientas informáticas que evalúan estadísticamente los resultados de las pruebas y ofrecen índices para valorar la bondad de los genes para este uso. Entre las herramientas más usadas se encuentran *geNorm*, *NormFinder* o *BestKeeper*. Por ejemplo, en el caso de *geNorm* se considera que un índice inferior a 1 para la expresión de un gen en las muestras analizadas, sería suficiente para que se pudiese usar como referencia en una RT-qPCR sobre esas mismas muestras.

En determinados casos, la elección de los genes a validar es complicada. Si se usa un criterio «clásico» se está condenado a emplear genes por tradición sin ningún apoyo experimental. Por ejemplo, realizar una búsqueda bibliográfica y emplear genes de referencia previamente publicados y con algún tipo de error en su uso o en el diseño de los oligonucleótidos. Un error muy común se debe a la nomenclatura de los genes empleados. En un organismo pueden existir diferentes parálogos (genes diferentes que producen un mismo tipo de proteína en un mismo organismo) y en los artículos ser todos nombrados de la misma forma. Esto provoca que, bajo una misma denominación, se usen diferentes genes para la normalización. Por ejemplo, bajo el nombre de *Actina* se usan genes diferentes en trabajos realizados en *Arabidopsis thaliana*, cuya expresión es diferente según los tejidos y condiciones experimentales. Además del uso de una correcta nomenclatura y de identificadores para solucionar este problema, se han ideado nuevas estrategias en la búsqueda de genes de referencia más allá de la tradición o de un concepto. La mayoría de ellas se basan en datos transcriptómicos a gran escala que evalúan la expresión de miles de genes en decenas, cientos o miles de muestras diferentes seleccionando aquellos genes cuya expresión sea más estable (menor desviación estándar).

A otro nivel, también es importante considerar un último aspecto sobre los genes de referencia y es que su validación se realiza basándose en una pareja de oligonucleótidos concreta. Es decir, que podemos diseñar parejas de oligonucleótidos para un gen de expresión estable que no sean adecuadas. Esto realza la importancia de seguir unas normas como las MIQE para el diseño de los oligonucleótidos que permitan una especificidad óptima con una alta eficiencia (Figura 1). En un artículo recientemente publicado por nuestro grupo de investigación<sup>[5]</sup> se comprueba cómo parejas de oligonucleótidos diseñadas bajo estos criterios presentan mejores resultados en cuanto a eficiencia y especificidad que las de otras parejas diseñadas sin tener en cuenta estos criterios.

Lo interesante de todo esto es comprobar cómo una técnica de rápida ejecución y que rinde gran cantidad de resultados es mucho más compleja de lo que los investigadores en general consideran. La bibliografía científica está repleta de pequeños y grandes problemas respecto a los resultados de la RT-qPCR. Y esto es si sólo tenemos en cuenta el efecto de una normalización deficiente de los resultados, en cuanto se profundiza algo más respecto a otros aspectos, como el diseño de las parejas de oligonucleótidos, el desastre se podría considerar mayúsculo. Y es que no hace más pan quien corre más sino quien trabaja mejor.

## Referencias

- <sup>1</sup>Huang T y otros. The mRNA of the Arabidopsis gene FT moves from leaf to shoot apex and induces flowering. *Science*. 309: 1694-1696, 2005.
- <sup>2</sup>Bustin SA y otros. The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. *Clin. Chem.* 55: 611-622, 2009.
- <sup>3</sup>Barber RD y otros. GAPDH as a housekeeping gene: analysis of GAPDH mRNA expression in a panel of 72 human tissues. *Physiol. Genomics*. 21:389-395, 2005.
- <sup>4</sup>Dheda K y otros. Validation of housekeeping genes for normalizing RNA expression in real-time PCR. *BioTechniques*. 37: 112-114, 2004.
- <sup>5</sup>Granados JM y otros. Selection and testing of reference genes for accurate RT-qPCR in adult needles and seedlings of maritime pine. *Tree Genet. Genomes*. 12: 60, 2016.

Diseño de oligonucleótidos

- Uso de regiones flanqueantes de intrones o de regiones de alta especificidad como 3 UTR
- Diseño de oligonucleótidos usando programas como Primer3:
  - Tamaños de 18-23 nt
  - Amplicones de 70-150 nt
  - Evitar índices de auto-complementariedad mayores de 0
  - Tm de los oligonucleótidos similares (menores de 3°C)
  - Contenido ideal de G+C de 50-60%

Verificación del plegamiento del amplicón

- Uso de programas como mFold para su verificación
  - La temperatura de "annealing" de la qPCR se considera la temperatura de plegamiento
  - Verificación de la ausencia de estructuras secundarias intrincadas
  - Descarte de la pareja de oligonucleótidos si el amplicón tiene estructuras con valores de  $\Delta G \leq -4$  kcal/mol

**Figura 1.** Ejemplo de criterios para el diseño de unos buenos oligonucleótidos para RT-qPCR. Figura propia basada en una anterior publicada por Granados JM y otros<sup>[5]</sup>.

y su *Diccionario de la lengua española* sirven de guía, pero, en realidad, son lo menos fiable para el lenguaje especializado.

Sugiero que cualquier profesor de ciencias tenga a mano el *Libro rojo* de Fernando A. Navarro (accesible por suscripción en [www.cosnautas.com](http://www.cosnautas.com)) y el *Diccionario de términos médicos* de la Real Academia Nacional de Medicina (recomiendo de nuevo la versión en línea y no el libro que se imprimió hace años). También se puede consultar mi pequeño grano de arena al respecto:

el libro *Ideas, reglas y consejos para traducir y redactar textos científicos en español*, cuya segunda edición está en prensa a cargo de la Fundación Dr. Antonio Esteve, mi Nanoblog, y las diferentes entradas que llevo escritas, y seguiré escribiendo, en esta revista.

Para los que estéis interesados en el tema, sabed que los próximos días 1, 2 y 3 de junio de 2017 se van a celebrar en Málaga unas jornadas sobre traducción científica y médica; podéis consultar el programa y cómo asistir en [www.asetrad.org/jtcm17](http://www.asetrad.org/jtcm17).

### Para saber más:

[Ideas, reglas y consejos para traducir y redactar textos científicos](#)  
[El nanoblog del Gonz](#)