

¿CÓMO FUNCIONA?

En busca del “ N_0 ”: siguiendo el camino de baldosas amarillas

Jan M. Ruijter[#], Adrián Ruiz Villalba*

[#]Dept Anatomy, Embryology and Physiology, Academic Medical Center
Amsterdam, The Netherlands
j.m.ruijter@amc.uva.nl

*Departamento de Biología Animal, Universidad de Málaga
adruiz@uma.es

En los últimos 15 años, la cuantificación de la reacción en cadena de la transcriptasa inversa (RT-qPCR) se ha convertido en el método más utilizado para la cuantificación de la expresión génica (Baker M, *Nature Methods*, 2011, 8(3): 207-212). Algunos de sus características han contribuido al uso tan extendido de esta metodología: i) es un ensayo en un solo paso, sin la necesidad de procesamiento post-PCR; ii) permite una comparación sencilla entre transcritos con un rango amplio de expresión (diferencias > 107 veces); y iii) utiliza el potencial cuantitativo implícito en una PCR convencional, haciéndolo un ensayo tanto cuantitativo como cualitativo (Bustin SA et al, *Journal of Molecular Endocrinology*, 2005, 34: 597-601).

Los principios básicos de este popular método de análisis son un conjunto de asunciones que tratan de solucionar el problema de la variabilidad asociada a cada uno de los pasos implicados en la RT-qPCR. El RNA se extrae, se estandariza y se retrotranscribe a cDNA. Se usan *primers* (cebadores) específicos para copiar y amplificar secuencias de interés en ciclos sucesivos de PCR hasta que la fluorescencia empleada para monitorizar los amplicones resultantes puede ser detectada. Finalmente, la fluorescencia observada por ciclo se usa para calcular la cantidad/concentración inicial del transcrito de interés; ésta se normaliza o calibra en función de un gen de referencia, caracterizado por mantener una expresión constante en las condiciones experimentales analizadas. Cada uno de estos pasos puede introducir sesgos y errores en el procedimiento (Baker M, *Nature Methods*, 2011, 8(3): 207-212). Para facilitar y viabilizar la comparación de resultados entre los diferentes experimentos y laboratorios, en 2009 se publicaron las *MIQE guidelines* como un intento de estandarizar todos los pasos implicados en la tecnología de la RT-qPCR (Bustin SA et al, *Clinical Che-*

mistry, 2009, 55(4): 611-622). Aquí se simplifican las 85 recomendaciones incluidas en dicha publicación en un diagrama que representa la RT-qPCR (figura 1). Las principales fuentes de variación y su posible solución se discuten paso a paso.

1. Diseño experimental

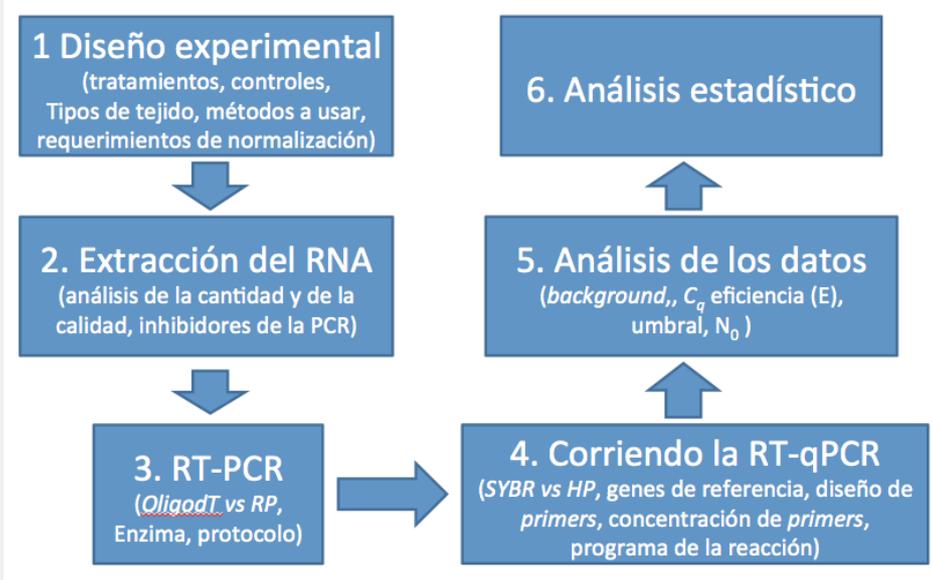
El diseño de experimentos biológicos es el paso más importante en el uso de la RT-qPCR. El tipo de tejido empleado y la forma de obtenerlo determinarán la calidad y la fiabilidad de la muestra y los requerimientos para la normalización de los resultados analizados. La estrategia de maximización de genes o de muestras se ha descrito como una estrategia de optimización para determinar eficientemente diferencias entre las muestras tratadas y los controles (Hellemans J et al, *Genome Biology*, 2007, 8:R19).

2. Extracción del RNA

El mercado ofrece diferentes métodos de extracción de RNA y los usuarios pueden elegir la opción más adecuada para su experimento (sales de guanidinio tipo Trizol, columnas de sílica-gel, *Cell-to-Ct* kits, entre otros). Tras la extracción, se analizan tanto la calidad como la cantidad de RNA usando métodos espectrofotométricos (*Nanodrop*, *2100 Bioanalyzer-Agilent Technologies*, *Experion-Biorad*). Tradicionalmente, la estabilidad del RNA ribosomal (rRNA) se ha considerado como una medida razonable para definir una extracción de RNA como óptima mediante un análisis en geles de electroforesis. Sin embargo, este punto de vista ha sido puesto en duda recientemente (Vermeulen et al, *Nucleic Acids Research*, 2011, 39(9):e63). La variabilidad asociada a este paso se puede

47

Figura 1. Esquema del flujo de trabajo de un experimento de la reacción en cadena de la transcriptasa inversa (RT-qPCR).



reducir usando el mismo método para todas las muestras analizadas.

3. RT-PCR

La mayoría de los usuarios obtienen cDNA a partir del RNA extraído y es en este paso donde se suele normalizar la cantidad de material de partida. Aquí se puede elegir entre diferentes primers para transformar el RNA en cDNA (oligo-dT, *random primers* o una mezcla entre ambos – Bustin et al, *Journal of Molecular Endocrinology*, 2005, 34: 597-601) al igual que entre diferentes protocolos de RT-PCR (Stahlberg A et al, *Clinical Chemistry*, 2004, 50(3): 509-515). La eficiencia de la RT-PCR es una fuente de variación muy importante en la tecnología de la RT-qPCR (Kitchen RR et al, *Methods*, 2010, 50(4): 231-236), ya que es clave asegurarse de que la enzima recorre todo el molde de mRNA, generando un set de transcritos representativo del transcriptoma de la muestra.

4. Corriendo la RT-qPCR

En este paso, el diseño/elección de *primers* (contenido en GC, secuencia del transcrito molde, longitud del amplicón); su optimización (temperaturas de *melting* y de *annealing*) y estandarización; el protocolo de la termocicladora; y la master mix que se va a usar son, entre otros aspectos, absolutamente esenciales. Existen muchas aplicaciones, como *Primer3Plus* o *RTPrimerDB*, ayudan al usuario a diseñar y estimar la combinación óptima de *primers*. La concentración de los *primers* afecta de forma muy directa a la línea base de fluorescencia (*baseline*) (Ruijter JM et al, *Nucleic Acids Research*, 2009, 37(6): e45), que es, como se verá más adelante, clave en la correcta estimación de los niveles de expresión del transcrito de interés. Otro de los factores que pueden introducir mayores fuentes de variación en el resultado final de la RT-qPCR es la elección de los genes de referencia. Normalmente, es necesario seleccionar genes de referencia para cada análisis, pero dado que hay ocasiones donde tanto económica como técnicamente esto no es posible, se recomienda el uso de al menos dos genes de referencia por ensayo (Vandesompele J, et al, *Genome Biology*, 2002, 3(7): research0034.1-0034.11).

5. Análisis de los datos de la qPCR

La cinética básica de la reacción de PCR se describe como $N_c = N_0 \times E^c$ donde la cantidad de amplicón en el ciclo c (N_c) es igual a la cantidad inicial de molde, N_0 , multiplicada por la eficiencia de la PCR (E) elevada al número de ciclos (c). Después de ajustar el umbral de fluorescencia (N_q) y determinar el número de ciclos necesario para alcanzar dicho umbral (C_q), la ecuación de la cinética puede ser invertida, lo cual permite el cálculo del N_0 a partir del umbral, de la C_q y de la eficiencia de la PCR (E): $N_0 = N_q / E^{C_q}$. Todas las ecuaciones generalmente empleadas en cuantificaciones tanto absolutas como relativas de expresión génica derivan de esta ecuación básica (ver tablas 1 y 2).

Hoy en día, la mayoría de las termocicladoras de qPCR del mercado estiman el parámetro C_q por cada reacción y amplicón. Este valor se calcula en relación al umbral elegido por el usuario (generalmente, el software de dichas máquinas permite ajustar un umbral de fluorescencia (N_q) tanto de forma manual como de forma automática). Con el valor de C_q obtenido usando una dilu-

48

TABLA 1. Ecuación básica de análisis de qPCR y cálculo de la expresión relativa del transcrito de interés.

$$\left. \begin{aligned} \text{Transcrito de interés: } N_{0,T} &= N_{q,T} / E_T^{C_{q,T}} \\ \text{Gen de referencia: } N_{0,R} &= N_{q,R} / E_R^{C_{q,R}} \end{aligned} \right\} \text{ratio de expresión por muestra}$$

$$\text{Ratio} = N_{0,T} / N_{0,R} = (N_{q,T} / E_T^{C_{q,T}}) (N_{q,R} / E_R^{C_{q,R}})$$

$$\text{Si } N_{q,T} = N_{q,R} = E_R^{C_{q,R}} / E_T^{C_{q,T}} = \begin{cases} \text{asumiendo } E_R = E_T \rightarrow E^{(C_{q,R} - C_{q,T})} \\ \text{asumiendo } E = 2 \rightarrow 2^{(C_{q,R} - C_{q,T})} \end{cases}$$

TABLA 2. Ecuaciones para calcular las diferencias de expresión entre las muestras control y las tratadas.

$X = \text{veces que difieren entre sí los niveles de expresión entre muestras control y tratadas}$

$$X = 2^{(C_{q,R} - C_{q,T})_{TREATED} / 2^{(C_{q,R} - C_{q,T})_{CONTROL}} \quad \text{Livak, Methods, 2001}$$

$$X = (E_R^{C_{q,R}} / E_T^{C_{q,T}})_{TREATED} / (E_R^{C_{q,R}} / E_T^{C_{q,T}})_{CONTROL}$$

$$X = E_T^{(C_{q,T,C} - C_{q,T,Tr})} / E_R^{(C_{q,R,C} - C_{q,R,Tr})} \quad \text{Pfaffl, NAR, 2001}$$

$$X = (N_{0,T} / N_{0,R})_{TREATED} / (N_{0,T} / N_{0,R})_{CONTROL} \quad \text{Ruijter, NAR, 2009}$$

ción seriada, también llamada curva estándar, el *software* o el usuario estiman la eficiencia de cada amplificación.

Para calcular las diferencias entre las condiciones control y experimentales, la mayoría de los usuarios usan la ecuación $2^{-\Delta\Delta C_q}$ (Tabla 2, línea superior; Livak K and Schmittgen TD, *Method*, 2001, 25:402-408) e ignoran la eficiencia real obtenida en la PCR, asumiendo que esta tiene un valor constante de 2. Otros usuarios usan una cuantificación relativa de la eficiencia (se denomina *eficiencia corregida*) (Tabla 2, segunda línea; Pfaffl M et al, *Nucleic Acids Research*, 2001, 29(9): 2003-7). Sin embargo, la cantidad de transcrito de interés por muestra y amplicón puede ser calculada como $N_0 = N_q/EC^q$ y normalizada con los niveles de expresión de los genes de referencia; los cálculos de las diferencias de expresión del transcrito de interés pueden ser estimados a partir de los valores de N_0 , lo cual evita la asunción de que la eficiencia de la PCR es constante (Tabla 2, línea de abajo; Ruijter JM et al, *Nucleic Acids Res*, 2009, 37(6): e45).

6. Análisis estadístico

Tras el cálculo de la expresión relativa de los genes por cada réplica experimental, la comparación estadística entre 2 condiciones biológicas distintas se realiza mediante un análisis de *t-Student* (para distribuciones normales) o un test de U de Mann-Whitney (para distribuciones no normales). Para comparar 3 grupos o más, se recomienda el uso del ANOVA, conjuntamente a un análisis post-hoc como el test de Tukey, o el test de Kruskal-Wallis. Para estudiar la dependencia entre genes y condiciones experimentales también se pueden realizar análisis de correlación y agrupaciones homogéneas.

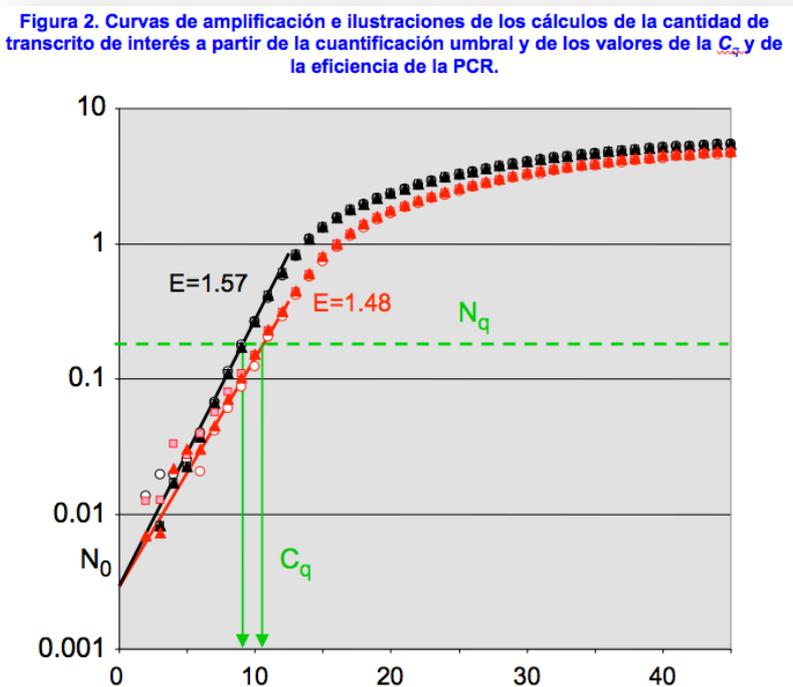
Con el objetivo de minimizar todas estas fuentes de variación potenciales, la mayoría de los usuarios usan el mismo método de extracción del RNA; fijan los criterios de calidad para el análisis de éste; emplean el mismo protocolo para la RT-PCR; y usan los mismos genes de referencia.

Las compañías que comercializan productos de RT-qPCR realizan un gran esfuerzo para facilitar el uso de esta tecnología a todos los usuarios y ofrecen muestras control, manuales de uso sencillos, soporte de problemas y dudas y organizan constantemente cursos o seminarios de formación para el diseño de este tipo de experimentos (Baker M, *Nature Methods*, 2011, 8(3): 207-212). A pesar de que se usan diferentes máquinas, aplicaciones y de que existe un mercado amplio de reactivos, toda la información dada tiende a facilitar la labor del usuario final para que éste solo tenga que *seguir el camino de baldosas amarillas*, que le llevará hasta el deseado *Mago de Oz*: el valor de la N_0 .

C_q , E y N_0

Como se ha comentado previamente, los valores de C_q , E y N_0 están estrechamente relacionados entre sí (figura 2). Sin embargo, se han descrito muchos métodos que estiman estos parámetros a partir de los datos de amplificación obtenidos. Es muy importante destacar como estas diferentes aproximaciones afectan al resultado final de expresión relativa del transcrito de interés en el análisis de la RT-qPCR.

La aplicación de las ecuaciones citadas en la figura 2 para el cálculo de N_0 desde N_q , C_q y E es matemáticamente idéntica al uso de la ecuación de la cinética de la PCR en el ciclo 0 de amplificación. Esto es más intuitivo cuando se representa la fluorescencia



Se muestran grupos de dos réplicas experimentales de curvas de amplificación donde la eficiencia de la PCR de una réplica (en rojo) fue inhibida al añadir *primers* modificados en 3' que previenen la elongación. La pérdida de eficiencia de la PCR conlleva un incremento del valor de la C_q . Los datos son obtenidos desde la base de datos *competimer* (<http://qPCRDataMethods.hfrc.nl>).

producida en un eje logarítmico: la región recta de la fase exponencial presenta una pendiente igual a $\log(E)$ y su extrapolación al ciclo 0 da el valor N_0 (figura 2).

La figura 2 muestra un experimento en el cual dos grupos de datos presentan la misma concentración del transcrito de interés, pero a uno de los grupos se le aplica una inhibición de la PCR (línea roja). Esto genera una eficiencia de amplificación menor, lo que conlleva un valor más alto de C_q . Sin embargo, al calcular el valor de N_0 con la ecuación $N_0 = N_q/E^{C_q}$, la diferencia del valor de la C_q es compensada por el 9% de diferencia en la eficiencia de la PCR, sin afectar, por tanto, a la cuantificación observada del transcrito de interés. Desde otro punto de vista, a pesar de las fuertes diferencias entre las eficiencias de la PCR, el investigador determina el valor correcto de N_0 en ambos grupos de réplicas. Así, si a la hora de indicar el valor de C_q se hubiese asumido una eficiencia constante, se hubiesen obtenido diferencias significativas NO REALES entre los dos grupos de análisis.

Este ejemplo muestra que es absolutamente esencial para el investigador estar seguro sobre la estimación de los parámetros de la RT-qPCR, especialmente del valor de la eficiencia de cada transcrito de interés. Los autores recomendamos ir un paso más allá: las *MIQE* indican cómo describir los valores de C_q y de la eficiencia, que los investigadores usarán principalmente como medida de la calidad del experimento. Sin embargo, el objetivo de un experimento de qPCR es determinar la cantidad del transcrito de interés, que se deriva del valor de la C_q individual tanto como de cada eficiencia de la PCR de cada transcrito. Describir sólo los valores de C_q , dejando fuera los de su eficiencia asociada, hace ignorar la dependencia de la qPCR a la eficiencia y puede llevar a resultados e interpretaciones erróneas. El valor de la eficiencia de la qPCR no solo es una propiedad del ensayo, sino un parámetro esencial en el cálculo correcto de la cantidad de transcrito de interés.

LinRegPCR: el camino de baldosas amarillas

Tal y como hemos demostrado previamente, la eficiencia de la qPCR es un parámetro esencial y muy sensible a la hora de obtener un valor correcto de N_0 . Se han publicado diferentes métodos que obtienen los parámetros de interés a partir de la curva de amplificación; todos estos métodos de análisis emplean diferentes algoritmos/aproximaciones para analizar las curvas de amplificación y determinar tanto la eficiencia como otros parámetros (Ruijter J et

al, *Methods*, 2013, 59(1): 32 - 46). Entre todos los parámetros que afectan fuertemente a la eficiencia, y en menor medida a la C_q , el más importante es la corrección de la línea de referencia de la fluorescencia (*baseline*) (Ruijter J et al *Nucleic Acids Res*, 2009, 37(6): e45). En base a esta premisa, nosotros recomendamos el uso del programa **LinRegPCR** (Ruijter J et al, *Nucleic Acids Res*, 2009, 37(6): e45) para analizar las curvas de amplificación de la RT-qPCR por tres razones principales:

1. LinRegPCR se basa en una sustracción de la línea base de fluorescencia que busca reconstruir la región lineal de la fase exponencial de una forma menos variable que la línea base dada por el sistema de la PCR después de que el usuario haya definido la fase de suelo de la reacción.

2. El programa define una ventana de datos que determina la mínima variación observada de la eficiencia de la PCR por amplificación y del C_q por reacción. Estos ajustes son automáticos, rápidos y fácilmente aplicables.

3. El programa genera pruebas de calidad basadas en el análisis de cada curva de amplificación. Adicionalmente, el usuario puede seleccionar manualmente los puntos más alejados de la media como consecuencia de un fallo durante la PCR en comparación con el resto de datos.

Estas tres propiedades del programa permiten al usuario obtener el valor real del N_0 por amplificación y por muestra de una forma muy intuitiva.

En resumen, concluimos que la determinación de la eficiencia de la PCR es al menos tan importante como determinar la C_q para llegar al valor real del N_0 . Determinar la eficiencia con el programa LinRegPCR es tan sencillo como seguir el camino de baldosas amarillas: buscar la fase exponencial de la curva de amplificación de cada transcrito de interés para obtener el valor real de N_0 .

