

**Empresa Brasileira de Pesquisa Agropecuária
Embrapa Caprinos e Ovinos
Ministério da Agricultura, Pecuária e
Abastecimento**

Documentos 112

On line

Considerações Estatísticas na Análise de Dados de Expressão Gênica Gerados pela Técnica de RT-qPCR

*Ana Maria Bezerra Oliveira Lobo
Raimundo Nonato Braga Lobo*

Exemplares desta publicação podem ser adquiridos na:

Embrapa Caprinos e Ovinos

Endereço: Estrada Sobral/Groaíras, Km 04 - Caixa Postal 145

CEP: 62010-970 - Sobral-CE

Fone: (0xx88) 3112-7400 - Fax: (0xx88) 3112-7455

Home page: <https://www.embrapa.br/caprinos-e-ovinos>

Sac: <https://www.embrapa.br/fale-conosco/sac/>

Comitê de Publicações da Unidade

Presidente: Francisco Selmo Fernandes Alves

Secretária-Executiva: Juliana Evangelista da Silva Rocha

Membros: Alexandre César Silva Marinho, Carlos José Mendes Vasconcelos, Diones Oliveira Santos, Maíra Vergne Dias, Manoel Everardo Pereira Mendes, Tânia Maria Chaves Campelo, Alexandre Weick Uchoa Monteiro e Viviane de Souza (Suplente).

Supervisor editorial: Alexandre César Silva Marinho

Revisor de texto: Carlos José Mendes Vasconcelos

Normalização bibliográfica: Tânia Maria Chaves Campelo

Editoração eletrônica: Comitê de Publicação

1ª edição on line (2014) - CGPE - 11526

Todos os direitos reservados

A reprodução não-autorizada desta publicação, no todo ou em parte, constitui violação dos direitos autorais (Lei n 9.610).

Dados Internacionais de Catalogação na Publicação (CIP)

Embrapa Caprinos e Ovinos

Lobo, Ana Maria Bezerra Oliveira.

Considerações estatísticas na análise de expressão gênica gerados pela técnica de RT-qPCR / Ana Maria Bezerra Oliveira Lobo e Raimundo Nonato Braga Lobo. – Dados eletrônicos. — Sobral: Embrapa Caprinos e Ovinos, 2014.

32 p. : il. (Documentos / Embrapa Caprinos e Ovinos, ISSN 1676-7659 ; 112).

Sistema requerido: Adobe Acrobat Reader.

Modo de acesso: < <http://www.cnpc.embrapa.br/publicacoes/>>.

1. Melhoramento Genético Animal. 2. Genética Animal. I. Lobo, Raimundo Nonato Braga. III. Título. V. Série.

CDD 21 ed. 576.5

© Embrapa 2014

Autoria

Ana Maria Bezerra Oliveira Lobo

Zootec., D. Sc., Pesquisadora da Embrapa
Caprinos e Ovinos. Fazenda Três Lagoas, Estrada
Sobral/ Groaíras, Km 04, Caixa Postal 145, CEP-
62010-970, Sobral/CE.

E-mail: ana.lobo@embrapa.br

Raimundo Nonato Braga Lobo

Méd. Vet., D. Sc., Pesquisador da Embrapa
Caprinos e Ovinos.

E-mail: raimundo.lobo@embrapa.br

Apresentação

A transcriptase reversa (RT) seguida de quantificação utilizando a técnica da reação em cadeia da polimerase (qPCR) em tempo real ou, simplesmente *reverse transcription-qPCR* (RT-qPCR), é um método *in vitro* para amplificação enzimática de sequências definidas de DNA complementar obtido a partir do RNA. A técnica de RT-qPCR é uma das mais utilizadas em experimentos de quantificação da expressão gênica por ser considerado um dos métodos mais sensível e específico de quantificação gênica dos que outros utilizados para esta finalidade.

No entanto, para obter resultados, pela técnica de RT-qPCR, que possam ser adequadamente analisados é necessário que o experimento tenha um desenho experimental com tratamentos, ensaios, unidade experimental e delineamento bem definidos. Muitas vezes, as considerações estatísticas, principalmente, no delineamento experimental para distribuição dos genes e das amostras (biológicas e técnicas) na placa de PCR (para controle de variação) não são consideradas e, dados gerados dessa maneira não são confiáveis. Após obtenção dos resultados da quantificação é primordial, antes de qualquer análise, avaliar a distribuição dos dados por meio de testes de normalidade.

Baseado nisso, recomenda-se, para a comunidade acadêmica, esta

Série Embrapa - Documentos como um roteiro prático para o correto delineamento experimental e a correta análise estatística dos dados para experimentos que utilizam a técnica de RT-qPCR.

Evandro Vasconcelos Holanda Jr.
Chefe Geral Embrapa Caprinos e Ovinos

Sumário

Introdução	09
Análises estatísticas de dados de qPCR em tempo real	11
Medida a ser utilizada na análise (Δ CT ou quantidade relativa $2^{-\Delta C_t}$)	12
Análise de normalidade e de homocedasticidade e transformação dos dados	21
Efeitos fixos e testes de significância a serem considerados na análise de dados de qPCR	27
Ajustes nos dados de qPCR para análise de correlação	31
Considerações finais	30
Referências	32
Literatura recomendada	33

Considerações Estatísticas na Análise de Dados de Expressão Gênica Gerados pela Técnica de RT-qPCR

Ana Maria Bezerra Oliveira Lôbo
Raimundo Nonato Braga Lobo

Introdução

A expressão gênica pode ser considerada como uma mediadora do fenótipo de um animal, uma vez que interpreta, por meio de uma molécula de RNA mensageiro (RNAm), as informações contidas no DNA das células e direciona a síntese de proteínas. Para quantificar a expressão gênica, Bustin (2000) sugeriu os seguintes métodos: *northern blotting* e hibridização *in situ* (PARKER; BARNES 1999), ensaios de proteção de RNase (HOD, 1992; SACCOMANNO et al., 1992), transcriptase reversa seguida por reação em cadeia da polimerase (WEIS et al., 1992) e arranjos de cDNA (*cDNA arrays* ou *microarrays*; (BUCHER, 1999). Há ainda os métodos de análise serial de expressão gênica ou SAGE (VELCULESCU et al., 1995) e o *differentially display* (ZHANG et al., 1998, LIANG 2002).

A transcriptase reversa (RT) seguida de quantificação utilizando a técnica da reação em cadeia da polimerase (qPCR) em tempo real ou, simplesmente *reverse transcription-qPCR* (RT-qPCR), é um método *in vitro* para amplificação enzimática de sequências definidas de DNA complementar (cDNA) obtido a partir do RNAm. Como o RNA não serve como molde para a PCR (PCR consiste na replicação *in vitro* de DNA), o primeiro passo realizado em um ensaio de RT-qPCR é a

transcrição reversa do RNA em cDNA (BUSTIN, 2000, p. 169). Dessa maneira, o método consiste na síntese de cDNA a partir de um RNA-molde, por meio da enzima transcriptase reversa, seguida pela amplificação, detecção e quantificação deste cDNA na reação da qPCR em tempo real. De acordo com Wang e Brown (1999), a qPCR pode ser um dos métodos mais sensíveis e específicos de quantificação gênica do que outros utilizados para esta finalidade.

Uma reação de PCR é caracterizada por três fases: a fase linear, a fase exponencial e a fase de platô. A PCR em tempo real explora o fato de que a quantidade de produtos de PCR na fase exponencial é uma proporção da quantidade inicial de *templates* (moléculas iniciais ou fitas molde) em condições ideais. Idealmente, durante a fase exponencial, os produtos de PCR irão duplicar durante cada ciclo, se a eficiência for perfeita, ou seja, 100%. É possível obter uma eficiência de amplificação por PCR próxima de 100% na fase exponencial se as condições de PCR, as características do *primer* (seguimento iniciador), a pureza do *template*, e os comprimentos de fragmentos amplificados forem ótimos. A dinâmica das fases da PCR é tipicamente observada através da ligação do DNA com compostos fluorescentes (SYBR® *Green*) ou pela hibridização com sondas, tais como *beacons* e TaqMan®. A emissão dos compostos fluorescentes gera um sinal que aumenta na proporção direta da quantidade de produto de PCR. Dessa maneira, a base da qPCR em tempo real é uma associação entre a fluorescência gerada com o número de amplicons (produto amplificado). Um gráfico da reação é gerado em tempo real (Figura 1). O gráfico refere-se ao sinal de fluorescência transformado pelo logaritmo de base 2 em relação ao número de ciclos e apresenta uma extensão linear, na qual o sinal da fluorescência logaritimizado correlaciona-se com a quantidade original de *templates*. Uma linha de base e uma de limiar pode ser ajustada para outras análises. O número do ciclo ao nível do limiar de fluorescência logaritimizado é definido como **cycle threshold** ou **Ct**, que é o valor observado na maioria dos experimentos de qPCR em tempo real, e, portanto, a principal mensuração de interesse estatístico. O Ct sumariza os resultados: quanto maior o

número de *template* no início da reação, um menor número de ciclos é necessário para alcançar um ponto, em que o sinal da fluorescência é primeiramente reconhecido como estatisticamente significativo, acima de um limiar arbitrário (*threshold*).

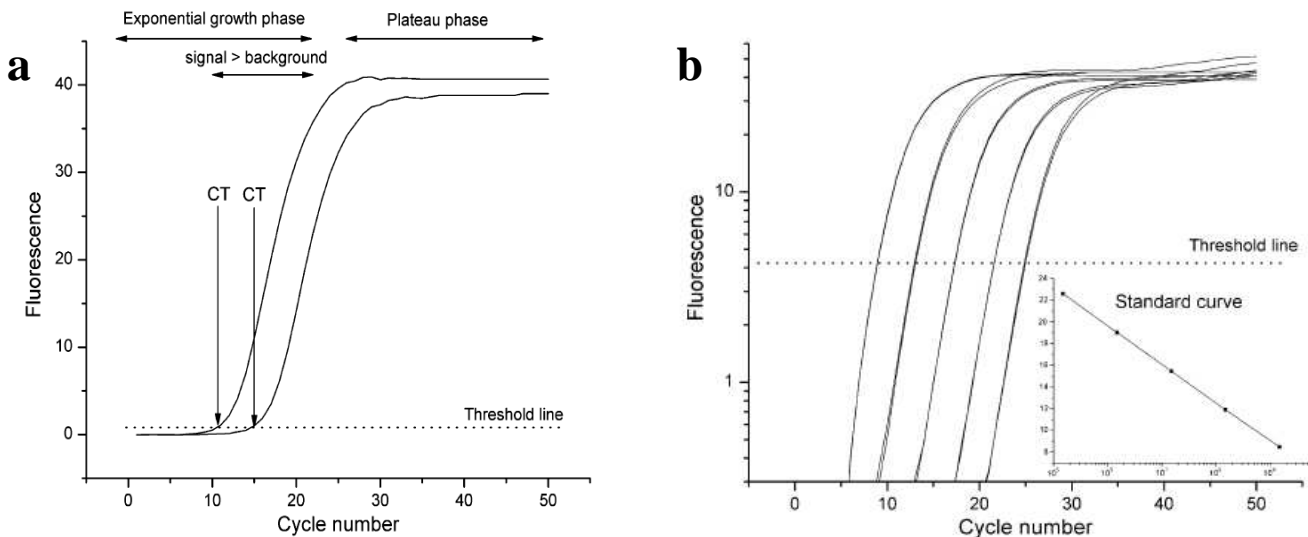


Figura 1. (a) Curvas da reação de PCR em tempo real. O número de ciclos (*cycle threshold* ou Ct) é detectado acima da linha de base (*threshold line*). (b) Escala logarítmica para cinco amostras padrões. O ponto de cruzamento com a linha de base são os valores de Ct. No detalhe, os valores de Ct são representados versus o logaritmo inicial do número de cópias do *template* na amostra padrão.

Fonte: Kubista et al. (2006).

Análises estatísticas de dados de qPCR em tempo real

Antes de qualquer execução de quantificação da expressão gênica, existe um longo caminho a ser percorrido para que se obtenham dados passíveis de serem analisados confiavelmente e que sejam capazes de confirmar ou não as pressuposições ou hipóteses a serem testadas. Portanto, um experimento para quantificação da expressão gênica envolve: a) desenho experimental: definir os tratamentos, o ensaio, a unidade experimental ou parcela e o delineamento experimental (inteiramente casualizado, delineamento em blocos casualizados, delineamento em quadrado latino, entre outros); b) coleta

de amostras de tecido ou células; c) extração do RNA e análise da qualidade e da sua quantidade; d) escolha dos genes-alvo e referência; e) desenho dos *primers* para os genes-alvo e referências; f) síntese do cDNA (transcrição reversa); g) teste de eficiência dos *primers* e análise da curva de dissociação; h) escolha dos *primers* baseado na eficiência; i) delineamento experimental para distribuição dos genes e das amostras (biológica e técnica) na placa de PCR (para controle de variação); j) realização da quantificação em tempo real ou qPCR; e, por fim, l) a análise dos dados (testes de normalidade e homocedasticidade, comparação de médias, análise de correlação, entre outras). Para os itens “b” ao “h” existem instruções que devem ser criteriosamente seguidas, conforme os guias: *The MIQUE Guidelines: Minimum Information for Publication of Quantitative Real-Time PCR Experiments* (BUSTIN et al., 2009) e o *Guideline to reference gene selection for quantitative real-time PCR* (RADONIC et al., 2004), e não é o objetivo deste documento discuti-las. Trataremos a seguir sobre o item l.

Medida a ser utilizada na análise (ΔCT ou quantidade relativa $2^{-\Delta Ct}$)

Quando se utiliza a técnica de RT-qPCR, empregam-se dois métodos de quantificação, a relativa e/ou a absoluta. Na quantificação relativa, os dados podem ser expressos em relação a um padrão interno (FINK et al., 1998), enquanto na absoluta os dados são expressos através de um curva padrão externa (BUSTIN, 2000), ou curva padrão (GINZINGER, 2002).

Na quantificação absoluta, os Cts obtidos a partir de uma amostra desconhecida são comparados aos Cts gerados a partir de uma série de amostras de concentração conhecida ou número de cópias. Segundo Giulietti et al. (2001, p. 391) para construir a curva padrão, utiliza-se amostra de concentração conhecida: um DNA dupla fita plasmidial contendo um inserto da região-alvo, RNA transcrito *in vitro*, DNA fita simples sintetizado *in vitro* (BUSTIN, 2000) ou qualquer amostra de cDNA expressando o gene-alvo. O número de cópias dessas amostras pode ser determinado por meio de espectrofotometria a 260 nm,

comparando-as a um peso molecular, seja DNA ou RNA (GIULIETTI et al., 2001). Dessa maneira, segundo Bustin e Nolan (2008, p. 37), os resultados podem ser expressos como número de cópias por unidade de massa, por exemplo, μg de RNA total.

A quantificação relativa não requer padrões com concentrações conhecidas. Nesse método, quantificam-se as amostras do gene-alvo em relação a um gene housekeeping ou controle endógeno (gene referência) e o número de cópias do gene-alvo são padronizados (ou normalizados) para o gene HK (ver Equação 1 a seguir). Na opinião de Bustin e Nolan (2013, p. 36),

“Em teoria, a quantificação relativa deve ser superior e muito mais conveniente do que a quantificação absoluta porque o resultado é uma razão, portanto, a concentração de RNA é irrelevante e há muitos modelos matemáticos que calculam a razão da expressão relativa, muita dos quais corrigem as diferenças de amplificação (LIU et al., 2002; PECCOUD et al., 1996; PFAFFL, 2001; PFAFFL et al., 2002), enquanto outros não (LIVAK et al., 2001)”

O modelo da calibração da eficiência (PFAFFL, 2001) e o modelo $\Delta\Delta\text{Ct}$ (LIVAK; SCHMITTIGEN, 2001) são os mais utilizados para a quantificação relativa. Segundo Yuan et al. (2006, p. 2), “o modelo da eficiência calibrada (Equação 1) é um modelo $\Delta\Delta\text{Ct}$ mais generalizado (Equação 2), sendo o sistema experimental para ambos os modelos similares”

$$\text{Razão} = \frac{(E_{\text{alvo}})^{\Delta\text{Ct}_{\text{alvo}}}}{(E_{\text{referência}})^{\Delta\text{Ct}_{\text{referência}}}}$$

Equação 1 (PFAFFL, 2001; YUAN et al., 2006)

Onde: $\Delta\text{Ct}_{\text{alvo}} = \text{Ct}_{\text{amostra controle}} - \text{Ct}_{\text{amostra tratamento (alvo)}}$ e, $\Delta\text{Ct}_{\text{referência}} = \text{Ct}_{\text{amostra controle (referência)}} - \text{Ct}_{\text{tratamento (referência)}}$

O modelo $\Delta\Delta\text{Ct}$ pode ser derivado do modelo acima se os genes-referência e o alvo atingirem a sua máxima eficiência de amplificação por PCR (Detalhes de como calcular a eficiência de amplificação são

descritos no ítem 2.1.1). Isso implica que a amplificação duplicou a cada ciclo, portanto, substituindo E na equação acima por 2 obtem-se:

$$2^{\Delta Ct_{alvo} - \Delta Ct_{referência}}$$

Yuan et al. (2006) definiu $\Delta\Delta Ct = \Delta Ct_{referência} - \Delta Ct_{alvo}$. Assim, a equação acima será:

$$2^{-(\Delta Ct_{referência} - \Delta Ct_{alvo})}$$

ou conforme apresentado abaixo:

$$Razão = 2^{-\Delta\Delta Ct}$$

Equação 2 (LIVAK; SCHMITTGEN, 2001).

Portanto, a quantidade de alvo (expressão relativa para um gene-alvo), normalizado para um controle endôgeno ou HK (ou gene-referência) e relativo a um calibrador, é então dada pelas equações acima demonstradas. Utilizando o método $2^{-\Delta\Delta Ct}$ o nível de expressão é obtido da seguinte maneira: os valores de Ct das amostras ou gene-alvo são subtraídos dos valores de Ct do gene-referência para obter os valores de ΔCt . Os valores de $\Delta\Delta Ct$ são obtidos pela subtração dos valores médios de ΔCt daqueles do calibrador de cada amostra. Demonstraremos a seguir como isso pode ser feito na prática. Neste exemplo, o objetivo é comparar se há diferença na expressão gênica entre raças (r1 e r2; N = 8, oito réplicas biológicas) para um gene-alvo (genA) relativo a um (1) gene referência (genR) utilizando três réplicas técnica (Tabela 1). Parte-se do princípio de que as amostras foram coletadas de animais submetidos a um experimento em delineamento inteiramente casualizado (ou seja, compunha o experimento animais contemporâneos e do mesmo sexo e sob a mesma dieta) e que as amostras foram coletadas e o RNA extraído seguindo todas as recomendações técnicas para obtenção de material de boa qualidade. O delineamento na placa de PCR deve ser feito com uma combinação de amostras técnicas e

biológicas. Para cada amostra biológica, são recomendadas três réplicas técnicas, tanto para o gene-alvo como para o gene-referência. Assim, cada réplica técnica para o gene-alvo terá a sua correspondente no gene-referência. Lembrando que em experimentos mais complexos, deve-se atentar para que todos os tratamentos sejam representados dentro da placa de PCR para reduzir variações. Quando isso não for possível, o efeito da placa deve ser considerado no modelo de análise dos dados. Conforme apresentado na Tabela 2, cada animal terá três Cts para o gene-alvo e três Cts para o gene-referência. O primeiro passo para a quantificação relativa é subtrair os Cts do gene-alvo daqueles do gene-referência para obter o ΔCt e calcular o ΔCt médio, embora muitos termocicladores já forneçam esse cálculo. Os valores médios são então aplicados na equação $2^{-\Delta\text{Ct}}$ e obtém-se, portanto, a quantificação relativa para cada animal para o gene de interesse (Tabela 2). Ressalta-se que o exemplo apresentado aqui é uma simplificação e, portanto, todos esses cálculos são feitos durante a análise estatística, sem necessidade de calculá-los a mão ou em planilhas de Excel. Por conseguinte, os valores de expressão relativa para os animais são comparados por meio de análise estatística, neste caso, compara-se as duas raças. Vale lembrar que anteriormente a realização deste ensaio deve-se proceder à análise de eficiência para se determinar quais os primers serão usados, bem como a concentração ideal de cDNA e os respectivos genes-referências para cada gene a ser quantificado (conforme será apresentado no próximo item).

Diante do apresentado até aqui, percebe-se que a principal mensuração de interesse na qPCR em tempo real é o número do ciclo ao nível do limiar ou Ct. Como já mencionado, o Ct é determinado de um gráfico log-linear do sinal da PCR versus o número de ciclo, sendo, portanto, um termo exponencial e não um termo linear (LIVAK; SCHMITTIGEN, 2001, p. 408). Segundo esses autores, “qualquer apresentação estatística usando os valores brutos de Ct deve ser evitada, à exceção é quando o interesse é examinar a variação intra-amostra entre reações replicadas”.

Tabela 1. Representação de um delineamento em uma placa de PCR de 96 “poços”.

	1	2	3	4	5	6	7	8	3	10	11	12
A	genA_an1r1	genA_an1r1	genA_an1r1	genA_an3r1	genA_an3r1	genA_an3r1	genA_an5r1	genA_an5r1	genA_an5r1	genA_an7r1	genA_an7r1	genA_an7r1
B	genR_an1r1	genR_an1r1	genR_an1r1	genR_an3r1	genR_an3r1	genR_an3r1	genR_an5r1	genR_an5r1	genR_an5r1	genR_an7r1	genR_an7r1	genR_an7r1
C	genA_an1r2	genA_an1r2	genA_an1r2	genA_an3r2	genA_an3r2	genA_an3r2	genA_an5r2	genA_an5r2	genA_an5r2	genA_an7r2	genA_an7r2	genA_an7r2
D	genR_an1r2	genR_an1r2	genR_an1r2	genR_an3r2	genR_an3r2	genR_an3r2	genR_an5r2	genR_an5r2	genR_an5r2	genR_an7r2	genR_an7r2	genR_an7r2
E	genA_an2r1	genA_an2r1	genA_an2r1	genA_an4r1	genA_an4r1	genA_an4r1	genA_an6r1	genA_an6r1	genA_an6r1	genA_an8r1	genA_an8r1	genA_an8r1
F	genR_an2r1	genR_an2r1	genR_an2r1	genR_an4r1	genR_an4r1	genR_an4r1	genR_an6r1	genR_an6r1	genR_an6r1	genR_an8r1	genR_an8r1	genR_an8r1
G	genA_an2r2	genA_an2r2	genA_an2r2	genA_an4r2	genA_an4r2	genA_an4r2	genA_an6r2	genA_an6r2	genA_an6r2	genA_an8r2	genA_an8r2	genA_an8r2
H	genR_an2r2	genR_an2r2	genR_an2r2	genR_an4r2	genR_an4r2	genR_an4r2	genR_an6r2	genR_an6r2	genR_an6r2	genR_an8r2	genR_an8r2	genR_an8r2

genA = gene alvo; geneR = gene referência; r1 = raça 1; r2= raça 2; an1 até an8 são as réplicas biológicas ou número de animais; todo animal ou réplica biológica e seu gene referência correspondente são analisados 3x (três réplicas técnicas).

Observação: A leitura da tabela deverá ser realizada considerando a análise de uma placa de PCR, ou seja contínua do 1 ao 12

Tabela 2. Representação simplificada dos cálculos para obter os Cts para o gene-alvo e o gene-referência. Estes cálculos são realizados para todas as réplicas representados na Tabela 1.

	Ct genA	Ct genR	ΔCt	ΔCt médio	$2^{-\Delta Ct}$
genR_an1r1 - genA_an1r1*	29	30	1	2	0,25
genR_an1r1 - genA_an1r1	28	30	2	-	-
genR_an1r1 - genA_an1r1	27	30	3	-	-
genR_an1r2 -genA_an1r2**	27,5	29,8	2,3	2,3	0,20
genR_an1r2 -genA_an1r2	28,3	30	1,7	-	-
genR_an1r2 -genA_an1r2	26,9	29,5	2,6	-	-

genA = gene alvo; genR = gene referência; r1 = raça 1; r2= raça 2; an1 = animal 1; *diferença entre as réplicas técnicas do Ct do gene-alvo (genA) e do gene referência (genR) para o animal 1 (an1) da raça 1 (r1); ** diferença entre as réplicas técnicas do Ct do gene-alvo (genA) e do gene referência (genR) para o animal 1 (an1) da raça 2 (r2).

Para examinar a variação intramostra, Livak e Schmittgen (2001) calcularam a média \pm desvio padrão (DP) para 96 amostras. A média \pm DP calculada do Ct bruto foi $20,0 \pm 0,194$ com um coeficiente de variação (CV) de 0,971%. Quando os valores individuais de Ct foram convertidos para a forma linear usando o termo 2^{-Ct} , a média \pm DP foi $9,08 \times 10^{-7} \pm 1,33 \times 10^{-7}$ com um CV de 13,5%. Assim, esses autores enfatizaram que os valores brutos de Ct falsamente representam a variação e devem ser evitados.

Pressuposições para o uso do método $2^{-\Delta\Delta Ct}$

Livak e Schmittgen (2011, p. 403) propuseram que, para o cálculo do $\Delta\Delta Ct$ ser válido, a eficiência de amplificação do gene-alvo e do gene-referência deve ser aproximadamente igual e sugeriram que, para se avaliar se dois *amplicons* possuem a mesma eficiência, é necessário observar como o ΔCt varia com as diluições seriadas dos *templates* ou cDNA (Figura 2).

Na prática, testa-se a eficiência de amplificação dos *primers* com as diluições seriadas de cDNA, normalmente na concentração de 1:10, 1:100 e 1:1000. No exemplo abaixo, (Tabela 3) demonstraremos a montagem de uma placa de PCR para a análise da eficiência de amplificação de três concentrações (100, 200 e 400 nM) para dois conjuntos de *primers* (p1 e p2) por gene (2) com três concentrações de cDNA (25, 75 e 225 ng/ul), sendo cada reação em duplicata. Esse procedimento

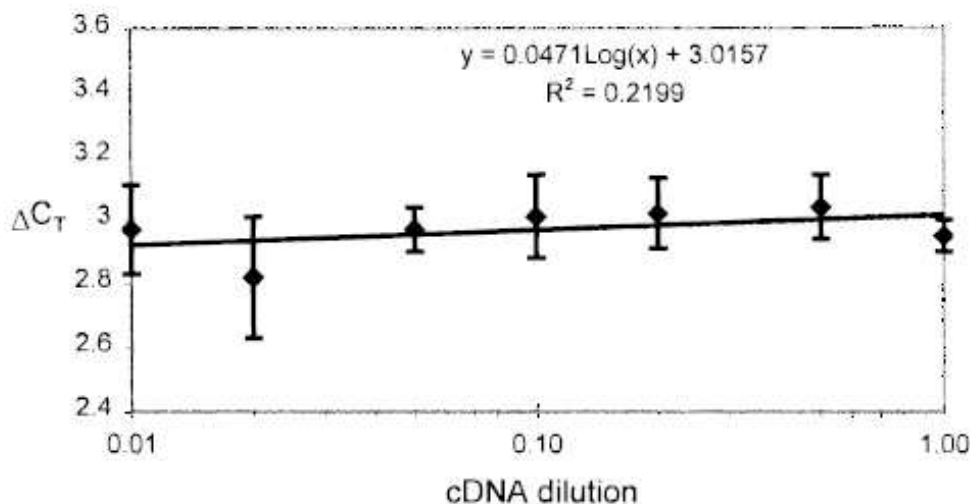


Figura 2. Validação do método 2^{-C_t} : amplificação de cDNA sintetizado de diferentes quantidades de RNA. A eficiência de amplificação do gene-alvo (c-myc) e do controle interno (referência, GAPDH) foi examinado usando PCR em tempo real e detecção por TaqMan.

Fonte: Livak e Schmittgen (2011, p. 403).

deve ser feito para todos os genes-alvo e os genes-referência que poderão ser utilizados posteriormente para a quantificação em si. Ressalta-se que em experimentos deste tipo, deve-se testar a amplificação de pelo menos dois conjuntos de *primers* por gene-alvo e testar de três a cinco genes-referência. Após a qPCR, de posse dos C_t s médio, calculados a partir das réplicas, procede-se a uma análise de regressão dos valores de C_t em função das concentrações do cDNA para obtenção do *slope* da equação. O valor do *slope* é aplicado posteriormente na equação da eficiência, conforme Tabela 4. Nessa Tabela, apresentamos os dados para um gene e um conjunto de *primers* e percebe-se que, conforme o valor do *slope* (-3,1506), o primer com melhor eficiência de amplificação é aquele na concentração de 100 nM. Dessa maneira, este valor de *slope* reflete em uma amplificação exponencial de 2,076 que indica que a reação duplicou a cada ciclo e, portanto, a reação de PCR teve uma eficiência de 1,07 ou 107% ($1,076 \times 100$; Tabela 4). Para comparações acuradas, em experimentos de qPCR, é necessário que a eficiência do gene-alvo e do gene-referência seja aproximadamente igual. Geralmente se aceita eficiência em torno de 90 a 110% e para isso o *slope* obtido deve corresponder a um valor entre -3,6 e -3,1 (Tabela 5), sendo o valor de -3,3 o ideal.

Tabela 3. Montagem prática de uma placa de PCR para a análise de eficiência de amplificação.

	100 nM			200 nM			400 nM			10	11	12
[cDNA]	1	2	3	4	5	6	7	8	9			
A	gen1-p1-25	gen1-p1-75	gen1-p1-225	gen1-p1-25	gen1-p1-75	gen1-p1-225	gen1-p1-25	gen1-p1-75	gen1-p1-225			
B	gen1-p1-25	gen1-p1-75	gen1-p1-225	gen1-p1-25	gen1-p1-75	gen1-p1-225	gen1-p1-25	gen1-p1-75	gen1-p1-225			
C	gen1-p2-25	gen1-p2-75	gen1-p2-225	gen1-p2-25	gen1-p2-75	gen1-p2-225	gen1-p2-25	gen1-p2-75	gen1-p2-225			
D	gen1-p2-25	gen1-p2-75	gen1-p2-225	gen1-p2-25	gen1-p2-75	gen1-p2-225	gen1-p2-25	gen1-p2-75	gen1-p2-225			
E	gen2-p1-25	gen2-p1-75	gen2-p1-225	gen2-p1-25	gen2-p1-75	gen2-p1-225	gen2-p1-25	gen2-p1-75	gen2-p1-225			
F	gen2-p1-25	gen2-p1-75	gen2-p1-225	gen2-p1-25	gen2-p1-75	gen2-p1-225	gen2-p1-25	gen2-p1-75	gen2-p1-225			
G	gen2-p2-25	gen2-p2-75	gen2-p2-225	gen2-p2-25	gen2-p2-75	gen2-p2-225	gen2-p2-25	gen2-p2-75	gen2-p2-225			
H	gen2-p2-25	gen2-p2-75	gen2-p2-225	gen2-p2-25	gen2-p2-75	gen2-p2-225	gen2-p2-25	gen2-p2-75	gen2-p2-225			

Observação: A leitura da tabela deverá ser realizada considerando a análise de uma placa de PCR, ou seja contínua do 1 ao 12.

Tabela 4. Cálculo da eficiência da amplificação exponencial e da eficiência da PCR para um dos genes representados na Tabela 3.

Gene	[] cDNA	[] primer	Ct médio	log[cDNA]	Slope $Y = a + bX + \varepsilon$	Amplificação Exponencial $E = 10^{(-1/\text{slope})}$	Eficiência da PCR $E = [10^{(-1/\text{slope})}] - 1$
gen1-p1	25	100	30,5905	1,3979			
gen1-p1	75	100	29,3265	1,8751			
gen1-p1	225	100	27,5840	2,3522	-3,150665153	2,076793409	1,076793409
gen1-p1	25	200	25,9560	1,3979			
gen1-p1	75	200	23,8430	1,8751			
gen1-p1	225	200	22,5275	2,3522	-3,592900049	1,898133247	0,898133247
gen1-p1	25	400	30,9475	1,3979			
gen1-p1	75	400	28,7590	1,8751			
gen1-p1	225	400	27,0055	2,3522	-4,131023071	1,746106727	0,746106727

Tabela 5. Variação do slope para uma eficiência entre 90% e 110%.

Slope	Amplificação	Eficiência
-3,6	1,8957	0,8957
-3,55	1,9129	0,9129
-3,5	1,9307	0,9307
-3,45	1,9492	0,9492
-3,4	1,9684	0,9684
-3,35	1,9884	0,9884
-3,3	2,0092	1,0092
-3,25	2,0309	1,0309
-3,2	2,0535	1,0535
-3,15	2,0771	1,0771
-3,1	2,1017	1,1017

Para um maior controle da qualidade dos dados de eficiência, Yuan et al. (2006) propõem ainda a análise de regressão linear após transformação logarítmica (\log_2) dos dados da concentração de cDNA. O modelo proposto é calculado usando o procedimento MIXED do pacote estatístico SAS (SAS, 2011). Assim, como na equação apresentada na Tabela 4, observa-se neste modelo uma relação linear entre o Ct e a concentração linearizada.

Análise de normalidade e de homocedasticidade e transformação dos dados

Para definir o método adequado de análise estatística dos dados de experimentos de RT-qPCR, é primordial que se conheça a distribuição e a natureza destes dados.

É sabido, portanto, que análises de (co) variância, entre outras, exigem distribuição normal dos dados e que eles possam ser parametrizados. As principais pressuposições a serem atendidas para a implantação de uma análise de (co) variância são a de normalidade dos dados e de variâncias iguais entre amostras de uma mesma população (homocedasticidade). O não cumprimento de uma dessas premissas

inviabiliza a análise, embora dados assim definidos possam ser analisados através de métodos não paramétricos ou por uma análise de variância após a transformação dos dados. Os testes não paramétricos são menos poderosos do que os testes paramétricos em que os dados seguem distribuição normal. Não obstante, os testes não paramétricos são projetados para ter propriedades estatísticas desejáveis, quando poucas pressuposições podem ser feitas sobre a distribuição dos dados.

Na análise de dados oriundos de experimentos de qPCR, uma primeira pergunta precisa ser respondida: os valores de Ct seguem distribuição normal?, Figura 3 (GONI et al., 2009). Existem vários métodos para verificação da condição de “normalidade” dos dados, entre eles, teste de qui-quadrado, análise de assimetria, análise da curtose, teste de Lilliefors, teste W: Shapiro-Wilk, teste Kolmogorov-Smirnov, etc., bem como, recursos gráficos, como histograma, Q-Q plot, P-P plot e normal plot. Essas estatísticas têm metodologia diferente para realização do teste de hipóteses. Assim sendo, faz-se necessário um estudo mais aprofundado de qual das metodologias é mais adequada. A análise de histograma é uma opção para a avaliação da normalidade, no entanto,

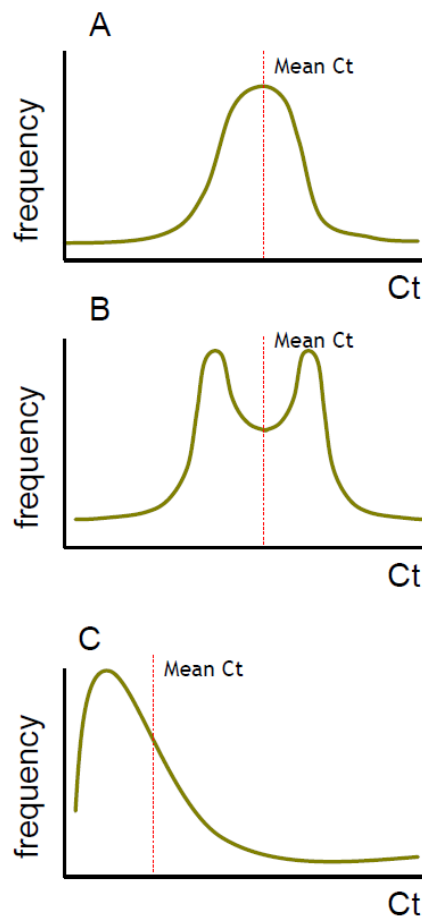


Figura 3. Possível distribuição dos valores de uma dada variável. Para uma determinada variável (ex.: expressão gênica) os possíveis valores (ex.: Cts) são representados no eixo de X, enquanto a frequência de tais valores é representada no eixo de Y. A representa uma distribuição normal; B representa uma distribuição simétrica, mas não com um único pico, portanto não normal; e C representa uma distribuição assimétrica, também não normal.

Fonte: Goni et al. (2009, p. 4)

não é uma medida quantitativa precisa de assimetria e curtose. O teste de Shapiro-Wilk é comumente utilizado quando o conjunto de observações é pequeno ($N \leq 50$), enquanto que o teste Kolmogorov-Smirnov é mais utilizado quando o conjunto de observações é grande ($N > 50$).

O teste de qui-quadrado produz resultados não confiáveis para pequeno tamanho de amostra. Q-Q e P-P plot apresentam a limitação de não gerar testes de significância para os resultados (LILLIEFORS, 1967). O teste de Kolmogorov-Smirnov baseia-se na máxima diferença entre a distribuição acumulada da amostra e a distribuição acumulada esperada. Quando a média e o desvio padrão da distribuição esperada não

são conhecidos, os valores de probabilidades de Lilliefors devem ser usados. Uma modificação foi feita por Lilliefors (1967), no teste de Kolmogorov-Smirnov para o seu uso quando a média e a variância não são especificadas, mas sim estudadas através dos dados da amostra (CECON et al., 2009, p. 149). As vantagens do uso do teste de Lilliefors em relação ao teste de qui-quadrado, por exemplo, são: i) pode ser aplicado sem restrição para pequenas amostras e ii) considera os dados individualmente, não perdendo informação devida a agrupamentos (CECON et al., 2009). O teste de Lilliefors pode ser realizado utilizando o pacote estatístico SAEG (SAEG 2007). Os testes de Kolmogorov-Smirnov e de Shapiro-Wilk podem ser executados utilizando o software R. O pacote estatístico SAS 9 (SAS, 2011) dispõe de vários testes para verificação da normalidade de um conjunto de dados, entre eles, Shapiro-Wilk, Kolmogorov-Smirnov, Anderson-Darling e Cramér-von Mises. Para maiores informações sobre esses testes, consulte o Guia de Procedimentos SAS (SAS, 2011). O procedi

são conhecidos, os valores de probabilidades de Lilliefors devem ser usados. Uma modificação foi feita por Lilliefors (1967), no teste de Kolmogorov-Smirnov para o seu uso quando a média e a variância não são especificadas, mas sim estudadas através dos dados da amostra (CECON et al., 2009, p. 149). As vantagens do uso do teste de Lilliefors em relação ao teste de qui-quadrado, por exemplo, são: i) pode ser aplicado sem restrição para pequenas amostras e ii) considera os dados individualmente, não perdendo informação devida a agrupamentos (CECON et al., 2009). O teste de Lilliefors pode ser realizado utilizando o pacote estatístico SAEG (SAEG 2007). Os testes de Kolmogorov-Smirnov e de Shapiro-Wilk podem ser executados utilizando o software R. O pacote estatístico SAS 9 (SAS, 2011) dispõe de vários testes para verificação da normalidade de um conjunto de dados, entre eles, Shapiro-Wilk, Kolmogorov-Smirnov, Anderson-Darling e Cramér-von Mises. Para maiores informações sobre esses testes, consulte o Guia de Procedimentos SAS (SAS, 2011). O procedi-

mento PROC UNIVARIATE pode ser utilizado para determinar se um conjunto de dados segue ou não distribuição normal. Esse procedimento gera uma variedade de sumários estatísticos, tais como, a média e a mediana, e representações numéricas das propriedades de skewness e curtose. Se a população a partir da qual os dados são obtidos é normal, a média e a mediana devem ser iguais ou muito próximas. O coeficiente de assimetria, que é uma medida de simetria, deve ser próximo de zero. Os valores positivos para o coeficiente de assimetria indicam que os dados são assimétricos à direita, e valores negativos indicam que os dados são assimétricos à esquerda. O coeficiente de curtose, que é uma medida de propagação, deve também ser próximo de zero. Os valores positivos para o coeficiente de curtose indicam que a distribuição dos dados é mais íngreme do que uma distribuição normal, e valores negativos para a curtose indicam que a distribuição dos dados é mais plana do que a distribuição normal. A opção NORMAL do PROC UNIVARIATE produz uma tabela com os testes de normalidade. Geralmente, os testes são realizados sob as seguintes hipóteses:

H_0 : a característica em estudo da população ou os erros (desvios) seguem a distribuição normal.

H_1 : a característica em estudo da população ou os erros (desvios) não seguem a distribuição normal.

Vale ressaltar, no entanto, que esses testes são muito dependentes do tamanho da amostra. Portanto, representações gráficas dos dados devem também sempre ser examinadas. Dessa maneira, devemos ter cautela no uso desses testes, devido as suas limitações. Ademais, experimentos de qPCR muitas vezes são realizados com baixo número de réplicas biológicas e isso dificulta a representação de normalidade ou não dos dados. Assim, no caso onde a normalidade é incerta, a análise não paramétrica pelo teste de Wilcoxon poderá ser usada (GONI et al., 2009).

Dados de qPCR além de não serem lineares, também apresentam heterogeneidade de variância entre as réplicas biológicas dentro e entre os tratamentos (RIEU; POWERS, 2009). Para a análise de homogeneidade de variâncias, os mais utilizados são o teste de Hartley, o teste de Cochran e o teste de Bartlett. O teste do F máximo de Hartley e o teste de Bartlett são utilizados para testar a homogeneidade das variâncias de duas ou mais populações. Segundo Cecon et al. (2009, p. 151), a eficiência do teste de Hartley é relativamente alta quando os números de observações são iguais e o número de grupos é ≤ 12 . As hipóteses a serem testadas, normalmente são:

H_0 : as variâncias entre as populações homogêneas, $\sigma_1^2 = \sigma_2^2 = \dots = \sigma_{t^*}^2$

H_1 : as variâncias entre as populações são heterogêneas.

Já o teste de Bartlett, é utilizado para comparar as variâncias de mais de duas populações, apresentando uma maior eficiência no julgamento das hipóteses, quando comparado ao teste de Hartley. Esse teste é aplicado quando as populações cujas homogeneidades serão testadas apresentarem distribuição normal.

O pacote estatístico SAEG (SAEG, 2007), dispõe de dois testes para análises de homogeneidade de variâncias. São eles: o teste de Cochran, para números iguais de repetição dentro de cada nível, e o teste de Bartlett, para variáveis com números desiguais de repetição. Se o conjunto de dados possui números iguais de repetição, o SAEG processará os dois testes.

Caso os dados originais provenientes do experimento de qPCR não apresentem distribuição normal ou apresentem heterocedasticidade, é possível realizar sua transformação para que essas pressuposições sejam atendidas e, assim, a análise estatística seja realizada.

Comumente, utiliza-se a transformação logarítmica dos valores de Ct, ou seja, $Ct' = \log_2(Ct)$. A transformação dos dados é realizada quando

se deseja inferir sobre os dados em sua escala original (CECON et al., 2009).

A transformação ou escalonamento (*range scaling*) ajusta a amplitude da distribuição a um novo intervalo de referência. Além da transformação logarítmica, há outras, tais como: a transformação angular ou arc sen e a radicial (para detalhes ver SAMPAIO, 1998). A transformação logarítmica é utilizada principalmente quando ocorre assimetria positiva, heterocedasticidade e o coeficiente de variação e a média são constantes (ou seja, o desvio padrão aumenta proporcionalmente ao aumento da média).

Deve-se atentar que não é apropriado realizar transformação e analisar os dados sem verificar se eles atingiram o objetivo, ou seja, passaram a apresentar distribuição normal após a transformação. Muitas vezes, após a transformação dos dados, é necessário aplicar novamente o teste de normalidade, bem como um teste de análise de homogeneidade de variâncias. Após esse procedimento, não atingindo a normalidade, os dados devem ser analisados usando um método não paramétrico mais adequado. Entre os principais testes não paramétricos, podem ser citados: Wilcoxon, Kruskal-Wallis, Friedman e Qui-Quadrado. O uso de cada um deles dependerá do tipo de experimento utilizado e da natureza da característica a ser analisada.

Efeitos fixos e testes de significância a serem considerados na análise de dados de qPCR

Em um experimento, diversas fontes de variação podem estar presentes, tanto aquelas que fazem parte do objetivo de estudo do pesquisador, como outras que não fazem parte desse objetivo, mas que não puderam, por diversos motivos, ser controladas. Por exemplo, um pesquisador deseja avaliar as diferenças de expressão de genes entre diferentes grupos genéticos, entretanto, havia animais de diferentes idades, diferentes manejos e de ambos os sexos. Seu objetivo é avaliar a diferença entre grupo genético (fonte de variação alvo), entretanto, será necessário avaliar os possíveis efeitos de idade, manejo, sexo,

que não puderam ser controlados. Além desses efeitos, dependendo do delineamento do experimento na placa de PCR, é necessário incluir no modelo de análise o efeito da placa.

Em experimentos cujos tratamentos estão distribuídos em uma única placa ou em delineamentos balanceados utilizando várias placas, nos quais todas as fontes de variação (tratamento, réplicas biológicas e técnicas) estão igualmente distribuídas dentro das placas, podem-se comparar os tratamentos por meio de uma análise de variância (ANOVA) dos dados transformados (utilizando o \log_2) e com distribuição normal (RIEU; POWERS, 2009). Nessa última situação, pode-se ainda controlar o efeito da variação entre placas, considerando-o no modelo como um bloco. A ANOVA não poderá ser usada quando utilizado um delineamento desbalanceado, ou seja, as fontes de variação não estejam igualmente distribuídas dentro das placas porque isto requer outra modelagem para estimar as médias dos tratamentos (RIEU; POWERS, 2009).

Na ANOVA considera-se tanto os efeitos-alvo como aqueles que não foram controlados. Efeitos classificatórios, não contínuos, entram na análise como efeito fixo (sexo, grupo genético, manejo, etc), enquanto aqueles com distribuição contínua, como pesos, idades, entre outras, são avaliados como covariáveis. Elas podem ser avaliadas com seus efeitos lineares e quadráticos, dependendo do caso.

Em experimentos onde os efeitos de sexo, manejo, idade, etc. não foram controlados, devem ser considerados no modelo e seus efeitos contabilizados. Portanto, o modelo de análise utilizado pelo pesquisador para comparar os dados de qPCR dependerá dos efeitos a serem testados e do delineamento experimental, da natureza da característica, entre outros.

Em modelos de análise de (co) variância (uma análise de covariância é aquela em que se têm efeitos fixos e covariáveis no mesmo modelo), os efeitos presentes são testados pelo teste F , que é um teste de razão

entre a variância da fonte a ser testada e a variância do resíduo. O valor dessa razão é comparado com o valor do teste F tabelado. Caso seja significativa, entende-se que essa variável está causando efeito significativo sob a característica em estudo e, assim, ela deve ser corrigida ou ajustada para esse efeito. Isso é importante, uma vez que, se os dados não forem ajustados para essas fontes de variação, a comparação entre os níveis da fonte de variação-alvo será viciada.

Assim, é importante avaliar diferentes modelos de análise, verificando o teste F das fontes de variação, bem como do coeficiente de determinação do modelo (R^2), para avaliar o mais adequado. O R^2 expressa o quanto da variância total foi explicado pelo modelo de análise. Ressalta-se que em alguns casos, mesmo que o teste F de alguma fonte de variação não seja significativo, essa fonte deve ser mantida no modelo, se sua presença promover redução na variância do resíduo e aumento do coeficiente de determinação.

Como o teste F é apenas um teste de razão de variância, ele não permite a verificação de diferenças entre diferentes níveis de uma fonte de variação, exceto quando essa fonte possui apenas dois níveis. Obviamente, nesse caso, um teste F significativo aponta diferenças entre esses dois níveis.

Com mais de dois níveis, é necessário, portanto, utilizar outros testes estatísticos como o t de Student, o SNK, o Tukey, o Scheffé, o Duncan, etc. A escolha do teste dependerá do controle dos erros experimentais (Erro Tipo I – atribuir significância quando ela realmente não existe e Erro Tipo II – atribuir equivalência quando realmente há diferenças entre os tratamentos), do número de níveis da fonte variação, bem como da instabilidade da característica analisada. Por outro lado, alguns testes controlam mais um tipo de erro e assim acabam por ser mais passíveis de cometer o outro tipo de erro. Além disso, a escolha do nível de significância (1%, 5%, etc.) poderá tender mais ou menos para um tipo de erro. A comparação das médias de uma fonte de variação neste caso está sendo realizada considerando a variância do resíduo, que foi

computada considerando todas as fontes de variação do modelo. Dessa forma, essa comparação é a mais adequada quanto mais for adequado o modelo de análise, com a presença de efeitos importantes. Essa comparação pode ser mais rigorosa, se em vez de analisar as médias simples da fonte de variação em questão, forem analisadas as suas médias dos quadrados mínimos (LSMEANS). Essas são médias ajustadas para todas as fontes de variação contidas no modelo. Assim, tanto a variância do resíduo foi controlada, como a média comparada foi corrigida para todos os possíveis efeitos presentes.

A escolha do nível de significância pelos autores que realizam experimentos de qPCR não está clara quando se deseja determinar se a expressão diferencial foi estatisticamente significativa ou se foi resultado do acaso. De acordo com Goni et al. (2009, p. 4), *“em geral, o padrão estatístico é que valores de p menores que 0.05 são considerados significativos (FISHER, 1925), porém, alguns autores atribuem um ponto de corte em nível de 0.01”*

São poucos os artigos de experimentos de qPCR que apresentam uma discussão aprofundada das considerações estatísticas nas análises para os efeitos de cada fator experimental, bem como não apresentam o valor de p e o intervalo de confiança (YUAN et al., 2006). Esses autores acreditam que “sem uma adequada modelagem e análise estatística, a interpretação de dados de PCR em tempo real pode conduzir o pesquisador a conclusões falso positivo” (YUAN et al. 2006, p. 3). Baseado nisso, esses autores desenvolveram quatro metodologias estatísticas para análise de dados de qPCR em tempo real, usando o método $\Delta\Delta Ct$ modificado: o modelo de Regressão Múltipla, Análise de Covariância (ANCOVA), teste t e teste de Wilcoxon (Tabela 6). O modelo de regressão múltipla considera os efeitos relacionados com o Ct , dentre eles a concentração de cDNA, o tratamento, o gene e suas interações. Esses autores concluíram que: o modelo de regressão múltipla e a ANCOVA foram similares por terem a mesma abordagem matemática para estimação do parâmetro. O teste t , embora apresente mesma estimação, fornece o erro padrão ligeira-

mente maior, conseqüentemente, haverá um grande intervalo de confiança. O teste de dois grupos de Wilcoxon conduz a estimação de um $\Delta\Delta Ct$ ligeiramente menor. Diante disso, os autores sugeriram que “a escolha dos modelos e/ou programas a serem utilizados irá depender do delineamento experimental e do rigor e da qualidade do experimento”.

Tabela 6. Comparação de quatro abordagens. Ct , erro padrão, valor de p e intervalo de confiança derivado de quatro métodos são apresentados.

Modelo	$\Delta\Delta Ct$	Erro padrão	Valor de p	Intervalo de confiança
Regressão múltipla	-0,6848	0,1185	<0,0001	(-0,4435, -0,9262)
ANCOVA	-0,6848	0,1185	<0,0001	(-0,4435, -0,9262)
Teste t	-0,6848	0,1303	<0,0001	(-0,4147, -0,955)
Teste de Wilcoxon	-0,6354	-	<0,0001	(-0,4227, -0,8805)

Fonte: Yuan et al. (2006).

Ajustes nos dados de qPCR para análise de correlação

Para se realizar análises de associação (correlação ou regressão) entre os valores de expressão gênica e características quantitativas produtivas de interesse, é necessário que estas estejam ajustadas para suas possíveis fontes de variação.

Assim, a característica em questão deve ser analisada na forma como foi descrita no item anterior, definindo todas as fontes de variação que a influenciam. Uma vez definido o modelo adequado, procede-se à estimativa dos valores preditos para cada observação. Esses valores preditos, que estão corrigidos para todas as fontes de variação no modelo, devem ser utilizados no estudo de associação (correlação).

Ressalta-se que as características em análise devem seguir distribuição normal multivariada. Em estudos de correlação em que essa pressuposição não é atendida, deve-se proceder à correlação de posto (rank) de Spearman, em vez da correlação de Pearson.

Considerações finais

Um dos primeiros passos para a realização da análise estatística deve ser investigar a distribuição dos dados. Muitos testes estatísticos dependem fortemente dos pressupostos da distribuição, tais como normalidade. Quando esses pressupostos não são satisfeitos, há uma maior chance de se cometer erro. A escolha do modelo adequado para análise estatística é dependente de questões científicas e do delineamento experimental. No entanto, análises estatísticas comparativas e valores de p parece ser a norma em dados de qPCR em tempo real. Autores recomendam teste de Wilcoxon quando se tratar de experimentos simples. Essa opção irá fornecer as diferenças significativas entre dois tipos amostrais e responder se as aparentes diferenças na expressão dos genes são estatisticamente diferentes.

Referências

BUCHER, P. Regulatory elements and expression profiles. **Current Opinion in Structural Biology**, v. 9, n. 3, p. 400-407, Jun., 1999.

BUSTIN, S. A. Absolute quantification of mRNA using real-time reverse transcription polymerase chain reaction assays. **Journal of Molecular Endocrinology**, v. 25, n. 2, p. 169-193, Oct., 2000.

BUSTIN, S. A.; BENES, V.; GARSON, J. A.; HELLEMANS, J.; HUGGETT, J.; KUBISTA, M.; MUELLER, R.; NOLAN, T.; PFAFFL, M. W.; SHIPLEY, G. L.; VANDESOMPELE, J.; WITTEWER, C. T. The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. **Clinical Chemistry**, v. 55, n. 4, p. 611-622, 2009.

BUSTIN, S. A.; NOLAN, T. Analysis of mRNA Expression by Real-Time PCR. In: SAUNDERS, N. A.; LEE, M. A. (Ed.). **Real-Time PCR: advanced technologies and applications**. Norfolk, Uk: Caister Academic Press, 2013. Cap. 3.

CECON, P. R.; SILVA, F. F. da; PETRNELLI, L. A.; OLIVEIRA, E. A. da; DETMANN, E. **Estatística experimental: conceitos, análises e interpretação**. Viçosa, MG: Universidade Federal de Viçosa, 2009. 220 p.

FINK, L.; SEEGER, W.; ERMERT, L.; HÄNZE, J.; STAHL, U.; GRIMMINGER, F.; KUMMER, W.; BOHLE, R. Real-time quantitative RT-PCR after laser-assisted cell picking. **Nature Medicine**, v. 4, n. 11, p. 1329-1333, Nov., 1998.

GINZINGER, D. G. Gene quantification using real-time quantitative PCR: An emerging technology hits the mainstream. **Experimental Hematology**, v. 30, n. 6, p. 503-512, Jun., 2002.

GIULIETTI, A.; OVERBERGH, L.; VALCKX, D.; DECALLONNE, B.; BOUILLON, R.; MATHIEU, C. An overview of real-time quantitative PCR:

applications to quantify cytokine gene expression. **Methods**, v. 25, p. 386-401, 2001. Disponível online em: <http://www.idealibrary.com>

GONI, R.; GARCÍA, P.; FOISSAC, S. The qPCR data statistical analysis. **Integromics White Paper**, Set. 2009. Disponível em: <<http://www.gene-quantification.de/integromics-qpcr-statistics-white-paper.pdf>>. Acesso em: 12 set. 2013.

HOD, Y. A simplified ribonuclease protection assay. **Biotechniques**, v. 13, n. 6, p. 852-854, Dec., 1992.

KUBISTA, M.; ANDRADE, J. M.; BENGTSSON, M.; FOROOTAN, A.; JONÁK, J.; LIND, K.; SINDELKA, R.; SJÖBACK, R.; SJÖGREEN, B.; STRÖMBOM, L.; STÅHLBERG, A.; ZORIC, N. The real-time polymerase chain reaction. **Molecular Aspects of Medicine**, v. 27, n. 2/3, p. 95-125, Feb., 2006.

LIANG, P. A decade of differential display. **Biotchniques**, v. 33, n. 2, p. 338-344, Aug., 2002.

LILLIEFORS, H. W. On the Kolmogorov-Smirnov test for normality with mean and variance unknown. **Journal of the American Statistical Association**, v. 62, n. 318, p.399-402, 1967.

LIVAK, K. J.; SCHMITTGEN, T. D. Analysis of relative gene expression data using real-time quantitative pcr and the 2ddct method. **Methods**, v. 25, n. 4, p. 402-408, Dec., 2001. Disponível em: <http://dingo.ucsf.edu/twiki/pub/Cores/BioinformaticsCore/QPCR/Livak_Schmittgen2001.pdf>. Acesso em: 12 set., 2013.

PARKER, R. M.; BARNES, N. M. mRNA: detection by in situ and Northern hybridization. **Methods in Molecular Biology**, v. 106, p. 247-283, 1999.

PFAFFL, M. W. A new mathematical model for relative quantification in

real-time RT-PCR. **Nucleic Acids Research**, v. 29, n. 9, p. 2002-2007, May, 2001.

RADONIC, A.; THULKE, S.; MACKAY, I. M.; LANDT, O.; SIEGERT, W.; NITSCHKE, A. Guideline to reference gene selection for a quantitative real-time PCR. **Biochemical and Biophysical Research Communications**, v. 313, n. 3, p. 856-862, 2004.

RIEU, I.; POWERS, S. J. Real-time quantitative RT-PCR: design, calculations, and statistics. **The Plant Cell**, v. 21, n. 4, p. 1031-1033, 2009.

SACCOMANNO, C. F.; BORDONARO, M.; CHEN, J. S.; NORDSTROM, J. L. A faster ribonuclease protection assay. **Biotechniques**, v. 13, n. 6, p. 846-850, Dec., 1992.

SAEG. **Sistema para análises estatísticas**, versão 9.1. Viçosa, MG: UFV; Fundação Arthur Bernardes, 2007. Disponível em: <<http://www.ufv.br/saeg/>>. Acesso em: 12 set., 2013.

SAS Institute Inc., System Requirements for SAS 9.2 Foundation for Microsoft Windows. Cary, NC: SAS Institute Inc., 2011.

VELCULESCU, V. E.; ZHANG, L.; VOGELSTEIN, B.; KINZLER, K. W. Serial analysis of gene expression. **Science**, v. 270, n. 5235, p. 484-487, Oct. 1995.

WANG, T.; BROWN, M. J. mRNA quantification by real time TaqMan polymerase chain reaction: validation and comparison with RNase protection. **Analytical Biochemistry**, v. 269, n. 1, p. 198-201, Apr. 1999.

WEIS, J. H.; TAN, S. S.; MARTIN, B. K.; WITTEWER, C. T. Detection of rare mRNAs via quantitative RT-PCR. **Trends in Genetics**, v. 8, n. 8, p. 263-264, Aug. 1992.

YUAN, J. S.; REED, A.; CHEN, F.; STEWART JUNIOR, C. N. Statistical analysis of real-time PCR data. **BMC Bioinformatics**, v. 7, p. 1-12, Feb., 2006.

ZHANG, J. S.; DUNCAN, E. L.; CHANG, A. C.; REDDEL, R.R. Differential display of mRNA. **Molecular Biotechnology**, v. 10, n. 2, p. 155-165, Oct. 1998.

Literatura recomendada

SAMPAIO, I. B. M. **Estatística aplicada à experimentação animal**. Belo Horizonte: Fundação de Ensino e Pesquisa em Medicina Veterinária e Zootecnia, 1998. 221 p.