

Real-Time Quantitative PCR as an Alternative to Southern Blot or Fluorescence *In Situ* Hybridization for Detection of Gene Copy Number Changes

Jasmien Hoebeek, Frank Speleman, and Jo Vandesompele

Summary

Changes in copy number of genes contribute to the pathogenesis of various genetic disorders and cancer. The status of a gene has not only diagnostic value but sometimes directs treatment stratification. Although, for many years, Southern blot and fluorescence *in situ* hybridization were the standard methods for the detection of deletion, duplication, or amplification of a gene, both methods have their own important limitations. Recently, real-time quantitative PCR has proven to be a good alternative for the detection of gene copy number changes. Its main advantages are the large dynamic range of accurate quantification, the absence of post-PCR manipulations, its high-throughput screening capacity and degree of automation, and the possibility to perform the assay on minimal amounts of sample DNA in just a few hours of time. In this chapter, we outline the procedure of how to develop an assay for the detection of gene copy number changes for your gene of interest. We illustrate the approach by describing a validated assay for the detection of germline *VHL* exon deletions and for determination of *MYCN* copy numbers in tumor samples.

1. Introduction

Alterations in gene copy number are responsible for many genetic disorders and lead to activation of oncogenes and inactivation of tumor suppressor genes in human cancers. Until recently, Southern blot and fluorescence *in situ* hybridization (FISH) were used as molecular detection methods for gene copy number changes. However, both methods have their own intrinsic limitations. Southern blot is laborious, time consuming, and requires large amounts of high-quality DNA. Although the strength of FISH is the direct visualization of the DNA copy number at the single-cell level, this method is not well suited for detection of partial gene deletions. In contrast, real-time quantitative PCR (qPCR) has evolved as a valuable

From: *Methods in Molecular Biology*, vol. 353
Protocols for Nucleic Acid Analysis by Nonradioactive Probes, Second Edition
Edited by: E. Hilario and J. Mackay © Humana Press Inc., Totowa, NJ

alternative in molecular diagnostics. This method has many advantages, including the absence of post-PCR manipulation (significantly reducing contamination risk and hands-on time), its speed, the requirement for only minimal amounts of input DNA, and the large dynamic range of accurate quantification.

The basic principle of real-time PCR is the monitoring of the PCR product accumulation at each PCR cycle during the reaction (in “real time”) as opposed to the end point detection by conventional (quantitative) PCR methods. PCR amplicon accumulation can be measured by a variety of fluorescent detection chemistries, such as the generic double-stranded DNA binding dye, SYBR Green I (**1**), or sequence-specific probes such as TaqMan[®] (**2**), adjacent hybridization probes (**1**), or molecular beacons (**3**). Many more detection formats are available; see the gene-quantification website section on dyes and chemistries (<http://dyes.gene-quantification.info>) and Chapter 17 for an overview.

Higuchi and colleagues (**4**) were the first to demonstrate that there is a relationship between the moment that the fluorescent PCR signal increases above background and the initial amount of template in the tube (**Fig. 1**). The sooner the signal increases above background, the more template was initially present. The difficult question of “how much?” is, therefore, transformed to a more easily addressed question of “when?” during real-time PCR analysis. In practice, by measuring fluorescence during PCR, a so-called amplification plot is generated. After background correction, an arbitrary fluorescence threshold line is set, such that it crosses the amplification plot in the exponential phase of the PCR (this is the linear phase in a semilogarithmic plot, and constitutes a range of PCR cycles in which the amplification efficiency is maximal, i.e., doubling of PCR product in each cycle under optimal conditions). The fractional PCR cycle at which the fluorescence crosses the threshold is called the threshold cycle (Ct) value. As stated, there is a linear relationship between this Ct value and the logarithm of the initial template amount. By measuring serial dilutions of a DNA sample, a standard curve can be constructed that allows intrapolation of the Ct value of an unknown sample to obtain relative or absolute quantities (depending on whether the actual copy numbers in the dilutions are known). In addition to the standard curve-based quantification strategy, another popular data analysis method (the comparative Ct method) is discussed in this chapter (see **Subheading 3.5.2.1**).

This chapter describes a real-time qPCR approach for the detection of gene copy number changes, such as deletions, duplications, or amplifications. The different steps include:

1. DNA isolation yielding PCR-grade DNA template.
2. Primer design (paying attention to primer specificity and avoiding amplicon formation with secondary structures) and experimental validation (evaluation on agarose gel [only once], melting curve analysis, and determination of efficiency).

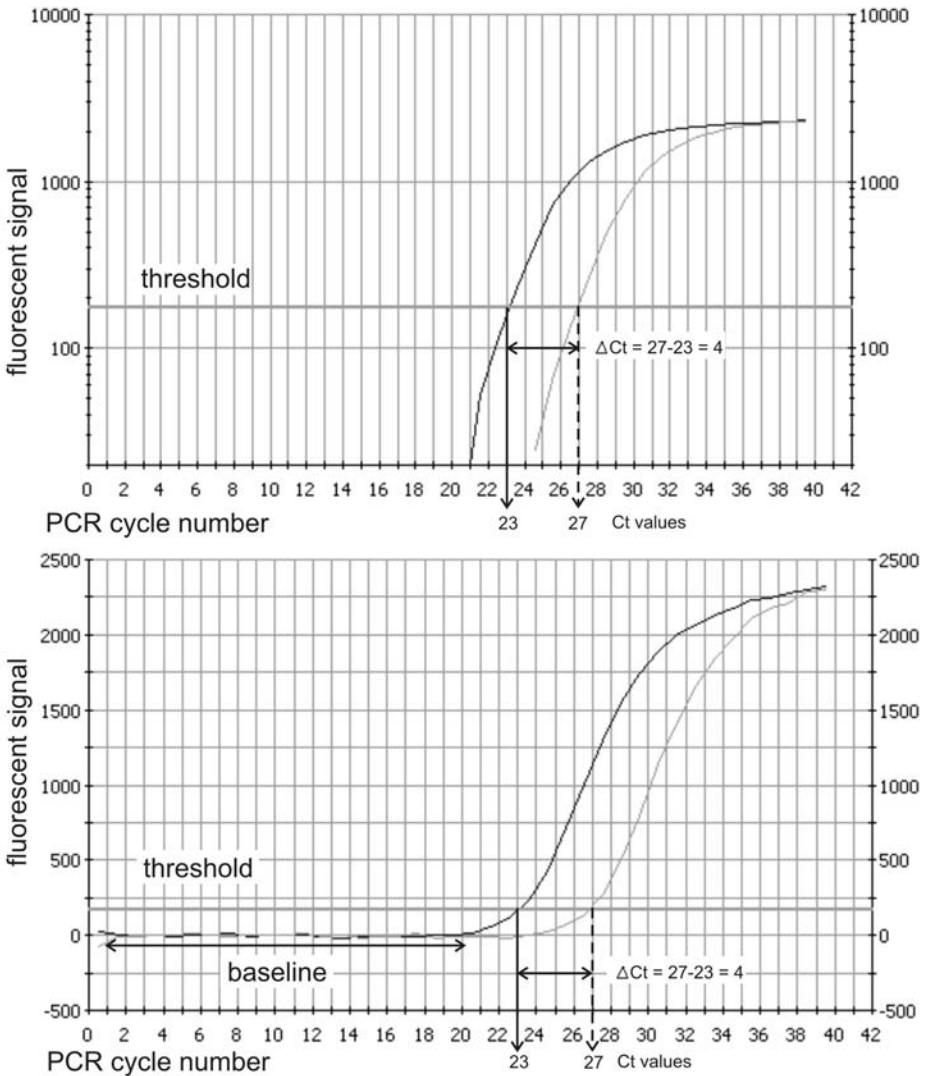


Fig. 1. Linear (bottom) and semi-logarithmic (top) amplification plot illustrating the nomenclature typically used in real-time PCR data analysis.

- 3. The actual PCR assay.
- 4. Data analysis.

In two different assays, we illustrate the use of real-time qPCR for gene copy number quantification, strongly demonstrating that qPCR is the method of choice for gene copy number assessment in routine DNA diagnosis. A first assay involves screening of germline *VHL* exon deletions in von Hippel-Lindau

(VHL) syndrome patients. The VHL syndrome is an autosomal-dominant inherited cancer syndrome caused by germline mutations or deletions of the *VHL* gene. In approximately one-fifth of all VHL syndrome families, small exon or larger gene deletions are detected. In a blind *VHL* exon deletion screening study by qPCR, we demonstrated that all deletions (exon or entire gene deletion in 10 of 24 samples) were detected. These data were in perfect agreement with previously determined Southern blot results, pointing at the high specificity and sensitivity of qPCR (5). A second assay was developed for measuring the *MYCN* gene copy number in neuroblastoma tumor samples. Amplification of the proto-oncogene, *MYCN*, is the hallmark of a genetic subgroup of aggressive neuroblastoma tumors and, as such, the status of the *MYCN* gene has become an important factor in therapy stratification. qPCR data for 175 neuroblastoma samples were highly concordant with previously determined FISH and Southern blot data (6).

2. Materials

1. PCR-grade DNA (fresh or frozen material; we have no experience yet with fixed paraffin-embedded material).
2. Primer design software (see **Subheading 3.3.3**).
3. Oligonucleotide primers: standard desalted, 250 μM stock solution in nuclease-free water and 5 μM working solution in nuclease-free water; store at -20°C (try to avoid more than 20 thaw–freeze cycles).
4. Real-time PCR master mix (see **Note 1**).
5. Real-time PCR instrument (see **Note 2**).
6. Optical 96-well reaction plate and optical adhesive sealing, or individual tubes with optical caps.
7. Filter tips for pipetting (aerosol barrier; see **Note 3**).
8. Microsoft Excel template for data analysis and visualization (available from the authors on request; see **Subheading 3.5.5**).

3. Methods

3.1. Template Preparation

Any protocol for DNA extraction that yields purified PCR-grade nucleic acids (removal of PCR inhibitors) will generate an adequate template for real-time PCR. We routinely use DNA from fresh or frozen cells (and have no experience with DNA from fixed paraffin-embedded tissue).

3.2. Reference Genes

Appropriate normalization of qPCR data is required to account for non-specific variation, such as variation in DNA input amounts or presence of PCR inhibitors, especially important in clinical samples. In the study of cancer, reference genes should be located in chromosomal regions that rarely show

genetic abnormalities in the cancer under investigation, i.e., *TNFRSF17* and *SDC4* in neuroblastoma, located on chromosome 16p13.1 and 20q12, respectively, which are not frequently altered in neuroblastoma (7). For other (constitutional) genetic disorders, in principle, every gene can be used as reference gene. In practice, reference genes are often selected in the vicinity or on the same chromosome on which the disease gene is located.

To validate candidate reference genes as proper reference genes, determine their copy number in both normal control DNA samples and affected patient or diseased tissue samples. The observed copy number ratio between the two reference genes in normal and patient DNA should be similar and both significantly equal to 1 (*see ref. 6* for more details).

3.3. Primer Design and Evaluation

The reliability and accuracy of real-time PCR lies, in part, in the specificity and reaction conditions of the primers and probe(s). Although many software tools and internet sites exist to design real-time primers and probes (summarized in **Subheading 3.3.3.**), validation experiments are required to ensure that the designed oligonucleotides perform well. It might prove difficult to design specific working primers and probes, particularly if working with highly homologous sequences (e.g., gene families).

As a golden rule, never use your old primer sequences (that you have been using successfully for many years in conventional PCR). You will save a lot of time and reagents by designing new primers, compatible with real-time PCR. Reaction condition optimization is almost never needed; follow the primer design (*see Subheading 3.3.2.*) and universal reaction setup guidelines (*see Subheading 3.4.*) and you have a high chance of a successful assay. Before designing your own set of primers, it might be interesting to take a look at the RTPrimerDB, a real-time PCR primer and probe database (8) (<http://medgen.ugent.be/rtprimerdb/>; limit search to “DNA copy number quantification/detection” applications) to see whether primers have been designed and validated for your specific gene of interest (GOI; *see Note 4*). We highly recommend submitting your validated primers to this database so that other people can benefit from your expertise. Furthermore, the parallel design of different primers for the same sequence in different laboratories is often a waste of resources and excludes a certain level of standardization and uniformity.

3.3.1. Sequence Retrieval and Preparation

Go to the Ensembl Genome Browser (<http://www.ensembl.org/>) and select your organism of interest. Search for the gene you are interested in and view (predicted) transcript information. In this frame, select “Exon information.”

Select and copy the sequence fragment of interest (*see Note 5*). Mask repeat sequences using RepeatMasker (<http://www.repeatmasker.org/>).

3.3.2. Primer Design Guidelines

Requirements for primer design:

1. Primer length: 9 bp/20 bp/30 bp (minimum/optimal/maximum).
2. Melting temperature (T_m) primers: 58°C/59°C/60°C (minimum/optimal/maximum).
3. GC content primers: 30%/50%/80% (minimum/optimal/maximum).
4. The five nucleotides at the 3'-end should have no more than two G and/or C bases, e.g., 5'-...ACTGT-3' (accept) vs 5'-...ACCGT-3' (reject). This above requirement can be set in the PrimerQuest software (*see Subheading 3.3.3.*) by adjusting the max 3' stability parameter to three.
5. Maximum T_m difference between primer pairs (°C): two.
6. Avoid runs of four or more identical nucleotides (especially G bases).
7. Amplicon length: 80 to 150 bp; avoid secondary structures by using the MFOLD software (*see Subheading 3.3.5.*).
8. If you use SYBR Green I as the detection chemistry, a probe is not needed.

3.3.3. Software

1. Primer3 (Whitehead Institute; free, Web-based [http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi]).
2. PrimerQuest (Integrated DNA Technologies; free, Web-based [<http://biotools.idtdna.com/primerquest/>]).

Many other dedicated software packages (free as well as commercial) are available (*see the “primer design” in the “links” section of the RTPimerDB, <http://medgen.ugent.be/rtpimerdb/>*).

3.3.4. BLAST Primer Specificity Analysis

Search for short nearly exact matches using a modified BLAST search at <http://www.ncbi.nlm.nih.gov/blast/> by pasting both primer sequences separated by 2 N bases (for further details, *see Fig. 2*).

3.3.5. MFOLD Secondary Structure Evaluation

For accurate quantification during real-time qPCR, a high PCR amplification efficiency is required. One of the main factors that may influence the efficiency of a PCR reaction, and, hence, the accurate detection of Ct values, is the formation of secondary structures in the amplified fragments in the region in which the primers (and probe, in case of probe-based detection chemistries) bind at the annealing temperature. Formation of a secondary structure is not favored and, hence, does not influence the PCR reaction if the hybridization of primer:target is thermodynamically stronger than the formation of a secondary structure. The

```
>gi|15431046|gb|AC069461.2|
```

```
D Homo sapiens chromosome 3 clone RP11-1022P15 map 3p, complete sequence
```

```
Length = 193214
```

1 Score = 56.0 bits (28), Expect = 2e-06

Identities = 28/28 (100%)

3 Strand = Plus / Minus

```
Query: 26      gctatcctgctacttaccacaacct 53
```

```
||||| |
```

```
Sbjct: 175800 gctatcctgctacttaccacaacct 175773
```

2 Score = 46.1 bits (23), Expect = 0.002

Identities = 23/23 (100%)

Strand = Plus / Plus

```
Query: 1       caatgctgacggacagcctatt 23
```

```
||||| |
```

```
Sbjct: 175700 caatgctgacggacagcctatt 175722
```

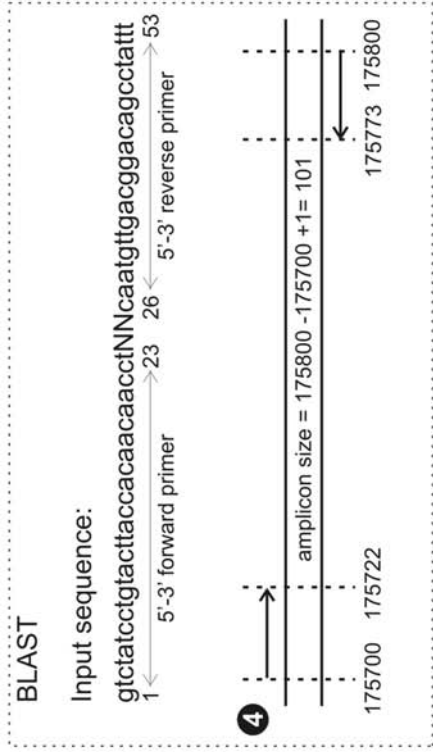


Fig. 2. Interpretation of BLAST output. The BLAST input sequence for a search of short, nearly exact matches against the chromosome or nonredundant database consists of the two primer sequences (5'-3'), separated by 2N bases (shown at the right side). Important BLAST result parameters are indicated in this figure: 1 "Expect" value describes the number of hits one can expect by chance if searching a database of a particular size. 2 "Identities" indicates the percentage of bases from the primer that correspond with the aligned sequence, the denominator of the fraction represents the primer length. 3 "Strand localization" the strand in the 5' to 3' direction of a DNA helix is called the "plus" strand, whereas the strand from 3' to 5' is the "minus" strand. 4 The PCR amplicon length can be indirectly calculated based on the information provided by the BLAST output (*see insert*). Good primers should meet the following requirements (for both forward and reverse primers):

1. An expect value close to zero; a cut-off cannot be proposed because this value depends on the primer length.
2. An identities value of 100% for both the forward and reverse primer.
3. Primers should be located on complementary strands.
4. An amplicon length between 80 to 150 bp (*see Subheading 3.3.2*).
5. Primers should only match via BLAST analysis at the sequence of interest. Primer pairs that result in one or more hits on the same or another chromosome are not specific.

Additional information can be found on the BLAST education site (<http://www.ncbi.nlm.nih.gov/Education/index.html>).

MFOLD v. 3.1 software (free, web-based), based on minimal free energy (<http://www.bioinfo.rpi.edu/applications/mfold/>; **ref. 9**) can be used for predicting the secondary structure of single-stranded DNA; applying default settings and a temperature that is the annealing temperature of the primers, 50 mM Na⁺, and adjusting concentration of Mg²⁺ to the actual concentration in the PCR reaction. Choose amplicons with minimal secondary structures. To determine which structures are favored, look at the ΔG (free energy; kcal/mole) value. The more negative the ΔG value, the stronger the formation of the structure is favored. **Figure 3** demonstrates that the presence of secondary structures in the region in which the primers anneal significantly influences PCR amplification efficiency.

3.3.6. Experimental Primer Evaluation

Always test a new primer pair by running a standard curve (*see* **Heading 1.** and **Subheading 3.5.2.2.**) and calculating the PCR reaction efficiency. Efficiency should be higher than 90% (slope > -3.6; **Fig. 3**). Run PCR samples once (and never again) on an agarose gel, to verify specificity and correct amplicon length (*see* **Note 6**). If using SYBR Green I, ensure the presence of a single sharp peak in the melting curve (*see* **Note 7**).

3.4. Reaction Setup

Provided that the same primer design guidelines are used, run all quantitative assays using the same universal thermal cycling parameters (10 min at 95°C [polymerase activation], 40 cycles at 95°C for 15 s [denaturation], and 60°C for 60 s [annealing and extension]; *see* **Note 8**). When working with SYBR Green I detection, start a melting curve run after the 40 amplification cycles. The described “universal” approach obviates any optimization of the thermal cycling parameters and means that multiple assays can be run on the same plate without sacrificing performance. The thermal cycling parameters constitute a two-step PCR (denaturation followed by combined annealing/extension), with an extra step for melting curve analysis if SYBR Green I is used as the detection format. Although initially recommended by some companies, real-time PCR reaction optimization (especially concentration of primers and MgCl₂) is now considered not to be required. Rigorous testing in our laboratory indicated that a final concentration of 250 nM for each primer is optimal for SYBR Green I assays. Many commercial ready-to-use master mixes or core kits for real-time PCR do not need any optimization, and provide accurate and reliable results (we have good experiences with the Bio-Rad and Eurogentec kits, working at 3 mM MgCl₂).

Table 1 shows a typical real-time PCR reagent mix with the individual reagents (*see* **Note 9**).

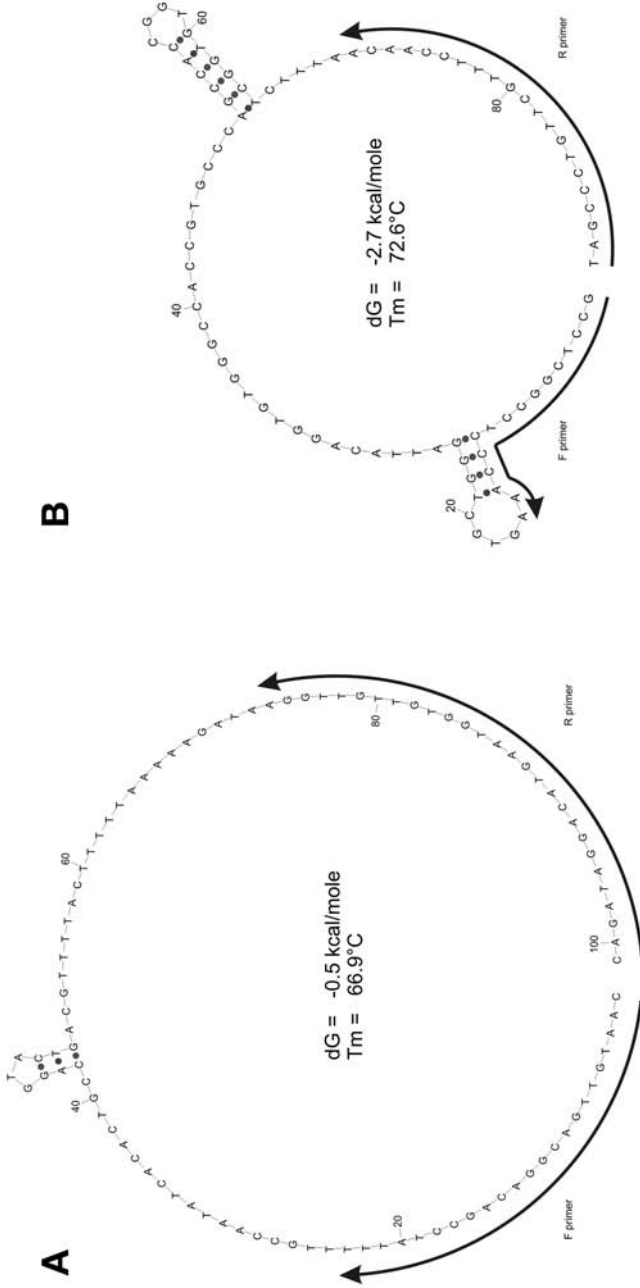
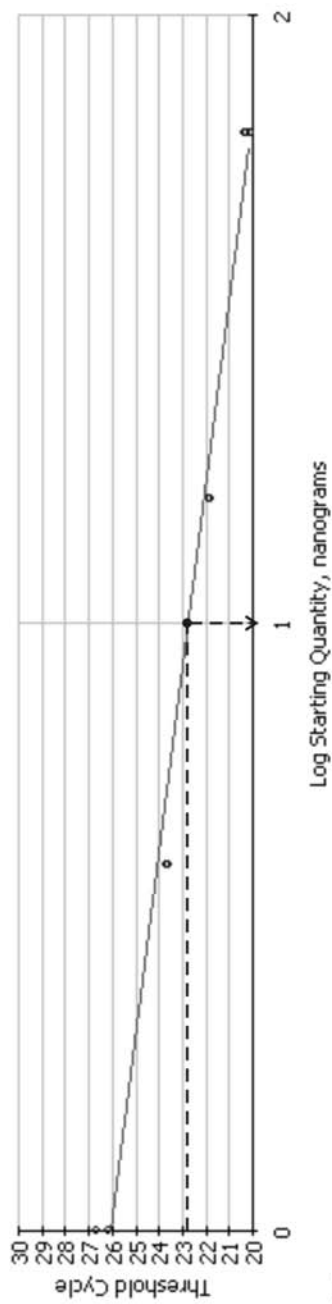


Fig. 3. (Continued)

C

Correlation Coefficient: 0.988 Slope: -3.296 Intercept: 26.042 $Y = -3.296 X + 26.042$
 PCR Efficiency: 101.1 %

● Unknowns
 ○ Standards

**D**

Correlation Coefficient: 0.909 Slope: -0.991 Intercept: 22.110 $Y = -0.991 X + 22.110$
 PCR Efficiency: 920.5 %

○ Standards

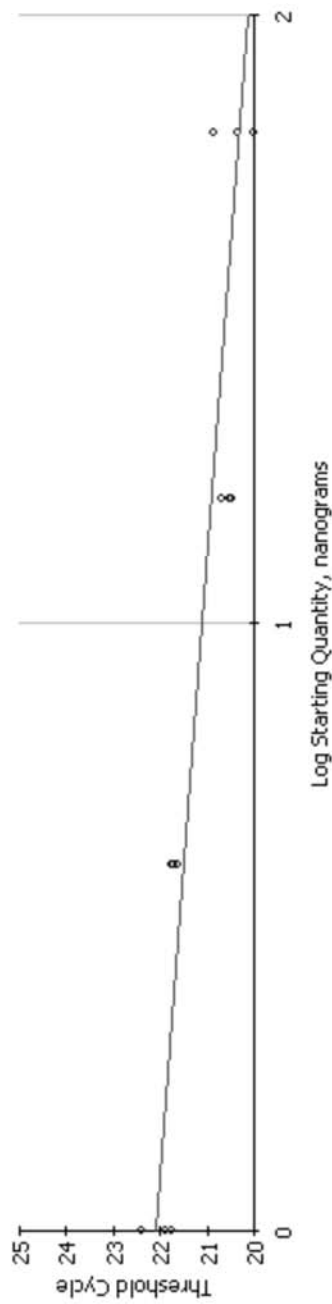


Table 1
Typical Real-Time PCR Reagent Mix

| | |
|-------------------------------------|-------------------------------------|
| SYBR dilution (1:2000) ^a | 0.75 μL |
| SYBR Green I PCR mix ^b | 12.5 μL |
| 5 μM Forward primer | 1.25 μL |
| 5 μM Reverse primer | 1.25 μL |
| 2.5 ng/ μL DNA template | 4 μL |
| Nuclease-free H ₂ O | to a 25- μL final volume |

^aDepending on the reagent kit, SYBR Green I might also be present in the PCR mix.

^bBuffer containing (hot-start) *Taq* polymerase, dNTPs (200 μM final concentration), stabilizers, and MgCl₂ (typically yielding 3 to 5 mM final concentration).

Refer to the instrument manual regarding whether a passive reference dye is required for calibration purposes (e.g., fluorescein for the Bio-Rad cyclers or ROX for the Applied Biosystems machines).

Always run duplicate reactions for gene amplification assays (example, *MYCN*) and quadruplicate reactions for deletion screening (example, *VHL*) for each sample, including the no-template control. Some practical hints for preparing a reaction plate are presented in **Notes 10–12**.

3.5. Data Analysis

3.5.1. Nomenclature

An amplification plot illustrating the nomenclature typically used in real-time PCR experiments is shown in **Fig. 1**. The amplification plot displays the fluorescence intensity vs the PCR cycle number. The baseline is defined as the PCR cycle range in which a signal is accumulating but lies beneath the detection limit of the instrument. The threshold line is used to define the Ct for each sample. The Ct is defined as the fractional PCR cycle number at which the fluorescent signal reaches the threshold value. Different methods are available to determine the threshold value. In one method, the threshold is calculated as the average baseline value plus 10 times the standard deviation (SD) of all baseline fluorescent signals. A fluorescent

Fig. 3. Secondary structure analysis and standard curves for calculation of PCR efficiency. (A) No secondary structures are present in the region where the primers anneal, the secondary structure between the primer annealing sites has a very small negative ΔG value, and, hence, does not influence the amplification efficiency (*see C*). (B) Secondary structure is more stable than primer:target hybridization and, hence, hampers efficient annealing of the primers (*see D*). (C) Using the primer set from (A), an almost perfect amplification efficiency (100%) is achieved. (D) The primer set from (B) results in an aberrant slope of -0.991 (920% efficiency), demonstrating the influence of secondary structures on PCR efficiency.

signal that is detected above the threshold is considered a real signal that can be used to define the Ct for a sample. Other methods arbitrarily position the threshold line in (the middle of) the linear part of the log-linear amplification plot (this is the first 5–10 cycles after the fluorescent signal increases above background, and indicates exponential amplification, in contrast to the later plateau phase). In practice, it does not matter very much where the threshold is set, as long as it is the same for all of the samples that you are comparing for your GOI.

3.5.2. Transformation of Ct Value to Quantity

3.5.2.1. COMPARATIVE CT METHOD

The Ct values of the different samples can be used to calculate the relative abundance of template for each sample. In **Fig. 1**, the solid line crosses the threshold at PCR cycle number 23, whereas the dotted line crosses at 27. By subtracting 23 from 27, there is a four-cycle difference between these two samples or a change in Ct (ΔCt) of 4. Because of the exponential nature of PCR, the ΔCt is converted to a relative abundance by $2^{\Delta\text{Ct}}$, or a 16-fold difference, in this case (*see Note 13*). This calculation forms the basis of the comparative Ct method for calculating DNA copy numbers. The ΔCt method generates raw (not normalized) quantities, which need to be normalized by dividing by a proper normalization factor (*see Subheading 3.5.4.*).

Another method that can be used to transform Ct values into normalized relative quantities is the $\Delta\Delta\text{Ct}$ method. This method relates the Ct value of your GOI in your sample to a calibrator/control sample and to the Ct value of a reference gene in both samples. Note that in the original publication of the $\Delta\Delta\text{Ct}$ method (**10**), there is no correction for a difference in amplification efficiency between the GOI and the reference gene (only the underlying requirement that the efficiency of both genes should be similar). Calculating $2 \times \Delta\Delta\text{Ct}$ between the GOI and the two reference genes, and taking the geometric mean of the two relative quantities is the same as first transforming the Ct values of the three genes into quantities using the ΔCt method, and dividing the GOI by the geometric mean of the two reference genes. Although both approaches yield the same result, we favor the ΔCt method, because:

1. It is much easier to perform in Excel.
2. It is very easy to take different amplification efficiencies for the different genes into account (just replace value 2 with the actual efficiency of the gene (e.g., 1.95 for 95%) in the formula of ΔCt).
3. It allows easy inclusion of multiple reference genes, which is a prerequisite for accurate normalization.

3.5.2.2. STANDARD CURVE METHOD

Besides the comparative Ct methods, a standard curve-based quantification method is also frequently applied for calculation of PCR efficiency and for

interpolating unknown sample quantities (**Fig. 3C**). Typically, a serial dilution of a positive control template is used to generate the standard curve. The resulting Ct values for each input amount of template are plotted as a function of the log concentration/quantities of input amounts (black circles), and a linear trend line is fit to the data. The resulting slope of the line fit to the data is used to determine the PCR efficiency, as shown in the formula. An ideal slope should be -3.32 for 100% PCR efficiency (i.e., a doubling of PCR product each cycle); in this example, it is 101.1%. The function that defines this slope is also used to calculate the amount of unknown samples by interpolation (black dot). Most real-time PCR instruments have software that can automatically compute the amount of template of an unknown sample from a standard curve. However, it can be calculated manually by putting the observed Ct value for an unknown sample into the formula: $(\text{observed Ct} - y \text{ intercept})/\text{slope}$.

3.5.2.3. CAVEATS IN DATA ANALYSIS WITH RESPECT TO REACTION EFFICIENCY

The main disadvantage of external standards (as described in **Subheading 3.5.2.2.**) is the lack of internal control for PCR inhibitors or other efficiency modulators in the DNA samples. All qPCR methods assume that the standards and the actual samples amplify with similar efficiency. Although this is true in most cases, this cannot be guaranteed in advance. The reaction efficiency calculated from the standard curve slope is, therefore, only valid for the standards, and is sometimes not transferable to the samples. The risk with external standards is that the unknown samples are amplified with varying efficiency. It is possible to test for PCR inhibitors directly to ascertain whether there is a consistent problem with unknown samples. Unrelated PCR product can be spiked into the sample, and the Ct value of the amplification can be compared with a control. A dilution series of the unknown sample can be tested, whereby PCR inhibitors are often diluted out, causing a nonlinear standard curve.

Although the standard curve is a valid quantification method in most cases, care should be taken to blindly extrapolate standard reaction efficiency to the unknown samples. Recently, several methods have been reported to determine the reaction efficiency for each individual tube based on the amplification plot. However, it is beyond the scope of this chapter to explain this in more detail. More info regarding this issue can be found on the GeneQuantification website (<http://efficiency.gene-quantification.info/>), where several approaches are listed.

3.5.3. DNA Melting Curve Analysis

An important issue related to the use of the cost-effective SYBR Green I dye is that it binds to any double-stranded DNA; the specific product, nonspecific products, and primer dimers are detected equally well. There are a number of ways to handle this problem. Careful primer design (including BLAST specificity

search, *see Subheadings 3.3.2. and 3.3.4.*), keeping primer concentrations relatively low (<300 nM), and increasing template concentration, can usually reduce primer dimers and nonspecific amplification. Hot-start techniques (in which the *Taq* polymerase is activated after an incubation at high temperature, usually several minutes at 95°C) is also helpful in reducing primer dimer and nonspecific amplification. The specificity of a SYBR Green I amplification can reliably and elegantly be monitored by generating a so-called melting curve of the PCR product (first described in **ref. 11**). Fluorescence is continuously monitored while heating the final PCR product from 60 to 95°C . Around the T_m of the amplicon (the temperature at which 50% of the products are denatured and single stranded), a sharp decrease in fluorescence is observed (because the SYBR Green I molecules can no longer bind the double-stranded DNA). With the instrument's software calculating the negative first derivative of the fluorescence over temperature, a typical melting peak is generated. A specific PCR product is characterized by a single sharp peak, whereas primer dimers generally display a lower and broader peak, at a lower temperature (**Fig. 4**; *see Note 14*). The resulting amplicon of a new primer pair is also analyzed using agarose gel electrophoresis (*see Subheading 3.3.6.*).

3.5.4. Accurate Normalization

Appropriate normalization is required to eliminate nonspecific variation, such as variation in DNA input amounts or presence of PCR inhibitors that are especially important in clinical samples. Using reference genes to normalize the data, it is not very important to use the same amount of DNA from each sample. However, for the different genes tested on the same sample, it is extremely important to use the same amount of the same concentration of DNA (and preferably originating from the same dilution). For screening copy number changes, two reference genes are needed. The use of two reference loci instead of one single locus for normalization of the data not only allows internal quality control, but also has proven to result in more robust, reliable, and accurate quantification (5).

3.5.5. Data Analysis and Error Propagation: Calculation of SD on Normalized Gene Copy Numbers

To calculate the SD on the final copy number of the normalized GOI (GOI_{norm}) copy number values, the error propagation rules for independent variables have to be applied.

3.5.5.1. SD OF RELATIVE COPY NUMBER VALUES

The ΔCt formula for transforming Ct values to copy numbers with the calibrator (normal control DNA sample) haploid copy number set to 1:

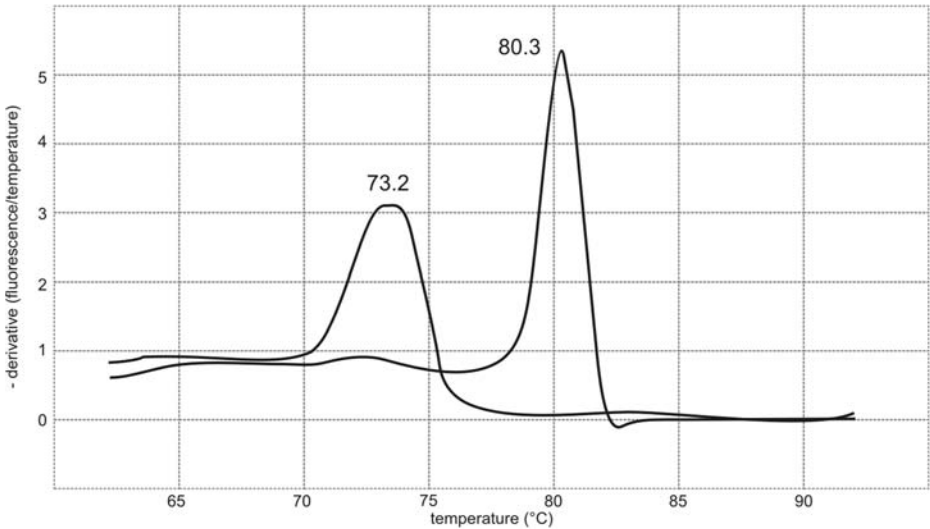


Fig. 4. Representative melting peaks of primer dimers and a specific amplicon, melting at 73.2 and 80.3°C, respectively. In general, primer dimer peaks are broader and smaller.

$$Q = E^{\Delta Ct} \tag{1}$$

$$Q = E^{(\text{calibratorCt} - \text{sampleCt})} \tag{2}$$

where Q is the sample quantity relative to the calibrator sample, E is the amplification efficiency (2 = 100%; see also **Note 13**), and calibratorCt is the Ct value of calibrator sample.

The SD for this relative quantity Q is (see geNorm manual for derivation of the formula, <http://medgen.ugent.be/genorm/>):

$$SD\ Q = E^{\Delta Ct} \times \ln E \times SD\ \text{sampleCt} \tag{3}$$

where lnE is the natural logarithm of the amplification efficiency and SD sampleCt is the SD Ct values of the sample replicates.

3.5.5.2. SD FOR NORMALIZED HAPLOID COPY NUMBERS

Two reference genes (REFG) and one GOI—each with their own SD values (calculated as outlined in **Subheading 3.5.5.1.**)—are measured, and the geometric mean of the two REFG is calculated as a reliable normalization factor (NF₂).

The normalization factor based on two reference genes is:

$$NF_2 = \sqrt{REFG_1 \cdot REFG_2} \text{ (geometric mean)} \tag{4}$$

The SD for this NF₂ is:

Table 2
Relative Quantities Calculated for the GOI and REFG

| | |
|------------------|----------------|
| GOI | GOI \pm SD |
| Reference gene 1 | REFG1 \pm SD |
| Reference gene 2 | REFG2 \pm SD |

GOI, gene of interest; SD, standard deviation; REFG, reference gene.

$$SD\ NF_2 = NF_2 \times \sqrt{\left(\frac{SD\ REFG_1}{2 \cdot REFG_1}\right)^2 + \left(\frac{SD\ REFG_2}{2 \cdot REFG_2}\right)^2} \quad (5)$$

The SD of the normalized GOI is:

$$SD\ GOI_{norm} = GOI_{norm} \times \sqrt{\left(\frac{SD\ NF_2}{NF_2}\right)^2 + \left(\frac{SD\ GOI}{GOI}\right)^2} \quad (6)$$

Alternatively, standard error (SE) values can be used instead of SD; the latter is the error on a single measured value and the former is the error on the mean (of repeated measurements).

$$SE = \frac{SD}{\sqrt{m}} \quad (7)$$

where m is the number of measurements, i.e., three for triplicates in a PCR. The error propagation rules are identical if using SD or SE values, just replace the SD values with SE values in **Eqs. 5** and **6**.

Note that this procedure only provides the error for the normalized gene copy number of a GOI for a single sample (mainly reflecting technical and pipetting variation, and variation among the different reference genes used for normalization).

However, if multiple samples are averaged (e.g., biological replicates [same cells independently grown or harvested], technical replicates [testing the same sample in different runs], or grouping samples with similar properties [e.g., diseased vs healthy tissue samples]), different rules apply (*see* the geNorm manual).

3.6. Example 1: Germline Exon Deletion Screening of the VHL Gene in Patients With VHL Syndrome

3.6.1. Template Preparation

Generate adequate template for real-time qPCR (*see Subheading 3.1.*). High-molecular-weight DNA was isolated from peripheral blood samples using the QIAamp DNA Blood mini kit according to the manufacturer's instructions (Qiagen).

3.6.2. Reference Genes

Select reference genes (*see Subheading 3.2.*). Two reference genes were selected on the same chromosome that the disease gene is located, i.e., *ZNF80* (3q13.2) and *GPR15* (3q12.1).

3.6.3. Primer Design and Evaluation

1. Examine the RTPrimerDB to check whether primers for your genes of interest are present. Primers for the *VHL*, *ZNF80*, and *GPR15* gene were not present in the RTPrimerDB (**8**).
2. If no suitable primer pairs are present in the database, design new primers following the designing guidelines (*see Subheading 3.3.2.*). We designed multiple primer pairs for each of the three *VHL* exons and for the two reference genes. Because of the small size of the *VHL* exons, we also developed primers spanning an exon–intron boundary (*see Note 5*).
3. Test primer specificity *in silico* (*see Subheading 3.3.4.*). Primer specificity was tested by a search for short nearly exact matches against the nonredundant and chromosome database, using the BLAST program.
4. Select primers with minimal secondary structures (*see Subheading 3.3.5.*). For the amplicons generated by the primers that passed the BLAST search, secondary structures of single-stranded DNA were predicted using the MFOLD program, applying default settings and 50 mM Na⁺, 3 mM Mg²⁺, and a temperature of 60°C (which is the annealing temperature of the primers). All primers that generate amplicons with secondary structures at the primer binding sites with a significant negative ΔG (free energy in kcal/mole) value were excluded from further analysis.
5. Validate your primers experimentally (*see Subheading 3.3.6.*). Amplification efficiency was evaluated by generating a standard curve. PCR products were run on an agarose gel to ensure correct amplicon length and specificity.
6. Primer sequences are available in the public RTPrimerDB database (<http://med-gen.ugent.be/rtpriimerdb/>; gene [RTPrimerDB-ID]: *VHL* [1023, 1024, and 1025], *GPR15* [1022], and *ZNF80* [1021]; **ref. 8**).

3.6.4. Reaction Setup (*see Subheading 3.4.*)

In addition to test samples, each assay must include:

1. A no-template control.
2. Calibrator DNA (a normal control or a mixture of genomic DNA from the organism of interest).
3. A positive control for deletion, duplication, or amplification (depending on the kind of assay to be developed).

We used the SYBR Green I detection method for quantification of the three test amplicons (*VHL* exon 1, 2, and 3) with unknown copy number and of the two reference genes *ZNF80* and *GPR15*. A 15 μ L reaction was performed following the cycling program: 10 minutes at 95°C, 40 cycling at 95°C for 15 s

and 60°C for 60 s. After PCR amplification, a melting curve is run to guarantee the specificity in every well of the reaction plate.

Each assay included:

- No-template control (in duplicate).
- 10 ng of calibrator DNA (in quadruplicate; Roche human genomic DNA).
- 10 ng of a positive control for deletion (in quadruplicate).
- 10 ng of test DNA (in quadruplicate).

3.6.5. Data Analysis and Error Propagation

1. First transform the Ct value quantities (either by using standard curves or the comparative Ct method). Here, calculation of the gene copy number was performed using the comparative Ct method (*see Subheading 3.6.2.1.*), according to the geNorm manual (<http://medgen.ugent.be/genorm/>; **ref. 12**).
2. Exclude reaction wells with obvious PCR reaction failure or clear outlier values (difference between Ct and mean Ct > 0.3) from further data analysis.
3. Calculate the arithmetic means of replicated Ct values.
4. Transform these arithmetic means to relative quantities with the copy number of the calibrator set to 1, using the Δ Ct formula, and an amplification efficiency of 97% (determined on the basis of a standard curve consisting of a DNA dilution series).
5. Normalize these relative quantities by dividing the *VHL* copy numbers by the geometric mean of two reference gene copy numbers (*ZNF80* and *GPR15*). Using this method, a haploid copy number of 1 is expected for a normal sample and a value of 0.5 for a sample with a *VHL* deletion.

We developed an Excel template for automated calculations, error propagation, and graphical representation of the results (available from the authors on request, **Fig. 5**).

3.7. Example 2: Detection of MYCN Oncogene Amplification in Neuroblastoma Tumors

Details regarding assay design and data analysis are described by De Preter and colleagues (**6**). Only substantial differences with the previous example are indicated in this section.

Two reference genes (*TNFRSF17* and *SDC4*) were selected in chromosomal regions that are rarely affected in neuroblastoma (based on comparative genomic hybridization results of more than 200 primary neuroblastomas; **ref. 7**).

Primer sequences are available in the public RTPrimerDB database (<http://medgen.ugent.be/rtpriimerdb/>; gene [RTPrimerDB-ID]: *MYCN* [11], *TNFRSF17* [14], and *SDC4* [15]; **ref. 8**).

Each assay included duplicate reactions instead of quadruplicate reactions because the difference in gene copy number caused by amplification is much larger compared with a single-copy loss in the case of a deletion.

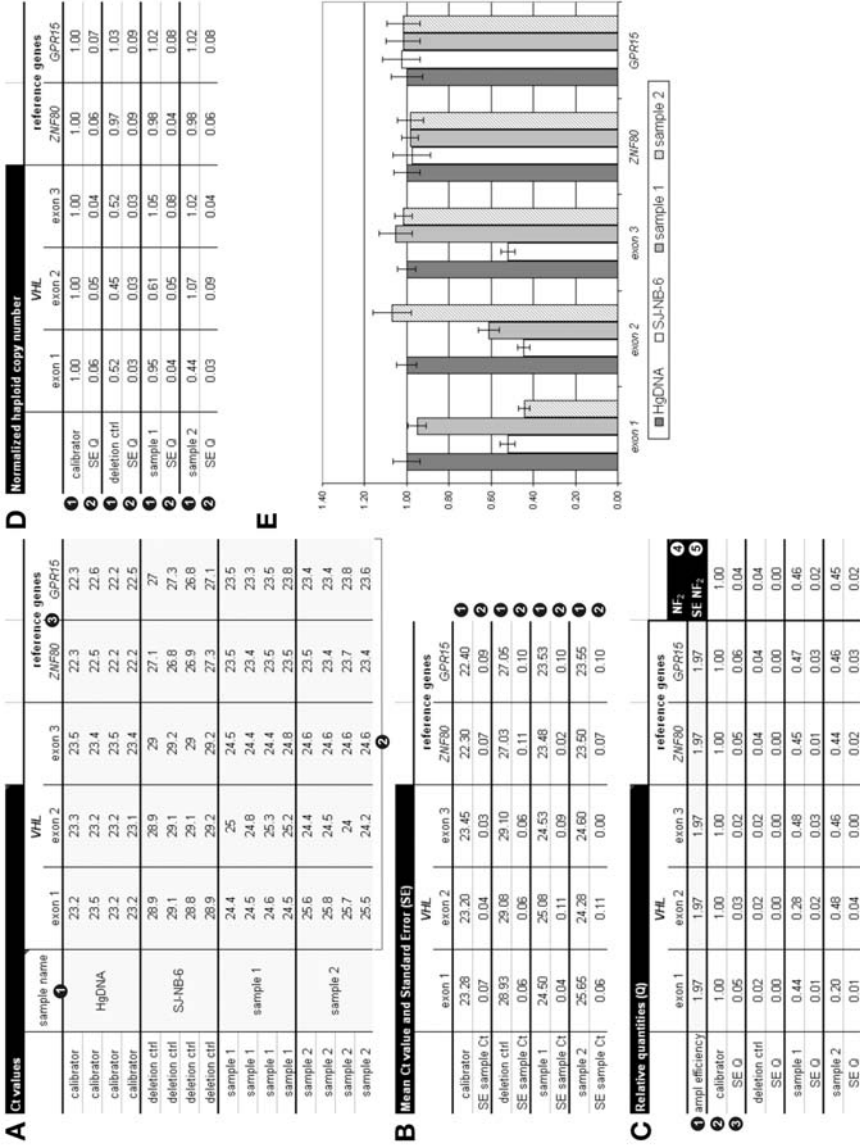


Fig. 5.

If using the standard curve method for data analysis in the case of gene amplification assays, use a dilution series of a sample that is amplified for the gene under investigation. In this way, the copy number values from both amplified and single-copy samples can be reliably interpolated from the standard curve.

3.8. Overview of Relevant Internet Links

1. Gene quantification: general website regarding all aspects of real-time PCR (quantification strategies, fluorescent detection chemistries, determination of amplification efficiency, and so on): <http://www.gene-quantification.info/>.
2. RTPrimerDB: public database for real-time PCR primer and probe sequences for all popular detection chemistries (with links to primer design software, real-time PCR machine vendors, and so on): <http://medgen.ugent.be/rtpimerdb/>.
3. geNorm: normalization using multiple reference genes: <http://medgen.ugent.be/genorm/>.
4. qpcrlistserver: qPCR discussion group: <http://groups.yahoo.com/group/qpcrlistserver/>.

4. Notes

1. We have had good experiences with iQ SYBR Green Supermix (Bio-Rad) or the SYBR Green I qPCR core reagents (Eurogentec).
2. Any real-time PCR machine is acceptable. Currently, we are using a GeneAmp 5700 thermal cycler (Applied Biosystems) and an iCycler iQ real-time PCR detection system (Bio-Rad).
3. A repetition pipet to distribute the master mix in the reaction tubes is recommended, especially if dealing with many samples, to increase reproducibility and to reduce hands-on time.
4. If you are new to real-time PCR, order a pair of primers (e.g., for a reference gene) that have been shown to work and convince yourself that you can perform the PCR and/or obtain good standard curves.

Fig. 5. (Continued) Microsoft Excel template for data analysis and visualization. (A) Importing data: ❶ Type sample name in this column. ❷ Paste Ct values for the three *VHL* exons and two reference genes for each sample (up to four replicates). ❸ Fill in reference gene names. (B) Calculation of mean Ct values and SE: ❶ In these rows: arithmetic mean of replicates. ❷ In these rows: $SE = SD/\sqrt{4}$ (in case of four replicates) (see **Subheading 3.6.5, Eq. 7**). (C) From Ct values to relative quantities: ❶ Paste amplification efficiency (see **Subheading 3.5.2.2**). ❷ Relative quantities (Q) are calculated using the comparative Ct method, using provided amplification efficiencies (C❶), and setting calibrator Q to 1. ❸ Error propagation using formula $SE\ Q = E^{ACt} \times \ln E \times SE\ sample\ Ct$. ❹ NF_2 , normalization factor based on two reference genes **Eq. 4**. ❺ Standard error NF_2 (based on Eqs. 5 and 7). (D) ❶ Normalization: Divide all relative quantities (Q) by the normalization factor NF_2 . ❷ Error propagation: $SE\ GOI_{norm}$ (based on **Eqs. 6 and 7**). (E) Graphical representation of the data.

5. If the exon you are interested in is too small or if no suitable primer pairs can be designed, you can add adjacent intron sequence and develop primers spanning an intron–exon boundary.
6. Some researchers advise sequencing the amplicon; however, this is not always straightforward, because of the small size of the fragment.
7. Sometimes a specific PCR product can generate a (atypical) melting curve with two or more peaks caused by regions in the amplicon that are characterized by different melting temperatures (for instance, a sequence with a GC-poor and GC-rich region).
8. Both annealing and extension can be performed at the same temperature (60°C); there is no need for a separate extension at 72°C during 1 min. Although the extension rate of the enzyme is slower at 60°C, it will certainly be sufficient to extend a 250-bp amplicon at 60°C for 1 min.
9. Having performed 25- μ L reactions for many years, we now routinely use a 15- μ L reaction. All components of the reaction mix are downscaled in the same proportion, but the amount of template DNA remains unchanged. In this way, we obtain a slight increase in sensitivity (lower Ct values, because of the higher initial template concentration) and a 40% decrease in reagent cost.
10. For the PCR master mix, make an excess of one reaction (for <20 reactions) or an excess of 5% (when dealing with >20 reactions). Always run duplicate reactions for each sample, including the no-template control. Work with filter tips in a dedicated PCR workstation (no flow) equipped with UV decontamination bulbs.
11. Prepare the reaction mix (reagents for quantification and primers) in a pre-PCR room to avoid carry-over contamination; this is a room different from the lab in which the DNA is prepared or in which post-PCR manipulations are performed. It might also help to use uracyl-*N*-glycosylase and dUTP nucleotides in the PCR. During an initial step at 50°C, the uracyl-*N*-glycosylase enzyme cleaves contaminating PCR products (carry-over from previous runs).
12. After preparing a 96-well plate for qPCR analysis (reaction mix and added DNA samples), shake the plate on a plate shaker (or vortex) to mix the DNA with the reaction mix, and centrifuge the plate shortly to spin down the reaction mixture and remove air bubbles. Always check the wells for air bubbles, because air bubbles can cause unusual reaction plots and, hence, inaccurate quantification of the gene copy number.
13. In the comparative Ct method, the value of 2 is used as the base in the formula $2^{\Delta Ct}$. A value of 2 means that the reaction efficiency of the PCR was 100%, which is almost never the case. The base value should be adjusted to the actual PCR efficiency (e.g., a value of 1.90 should be used if the efficiency is 90%). Most people determine the reaction efficiency once using a standard curve, and, later on, use this value in their comparative Ct analytical procedure.
14. Although it is possible to collect fluorescence data at a higher temperature (between the T_m of the nonspecific/primer dimer signal and the true signal; thus, removing the contribution of nonspecific signal to the measurement), this is not recommended because the simultaneous amplification of the nonspecific product(s) can adversely affect the amplification of your sequence of interest.

Acknowledgments

This work was supported by GOA-grant 12051397, FWO-grant G.0185.04, and VEO project 011V1302. We thank Bio-Rad Belgium for their financial and technical support. We gratefully acknowledge Katleen De Preter for her work on *MYCN* copy number determination and Els de Smet and Nurten Yigit for technical help with the qPCR experiments. Jo Vandesompele is supported by a grant from the Flemish Institute for the Promotion of Innovation by Science and Technology in Flanders (IWT).

References

1. Wittwer, C. T., Herrmann, M. G., Moss, A. A., and Rasmussen, R. P. (1997) Continuous fluorescence monitoring of rapid cycle DNA amplification. *Biotechniques* **22**, 130–131, 134–138.
2. Livak, K. J., Flood, S. J., Marmaro, J., Giusti, W., and Deetz, K. (1995) Oligonucleotides with fluorescent dyes at opposite ends provide a quenched probe system useful for detecting PCR product and nucleic acid hybridization. *PCR Meth. Appl.* **4**, 357–362.
3. Tyagi, S. and Kramer, F. R. (1996) Molecular beacons: probes that fluoresce upon hybridization. *Nat. Biotechnol.* **14**, 303–308.
4. Higuchi, R., Fockler, C., Dollinger, G., and Watson, R. (1993) Kinetic PCR analysis: real-time monitoring of DNA amplification reactions. *Biotechnology (NY)* **11**, 1026–1030.
5. Hoebeek, J., van der Lijjt, R., Poppe, B., et al. (2006) Rapid detection of VHL exon deletions using real-time quantitative PCR. *Lab Invest.* **85**, 24–33.
6. De Preter, K., Speleman, F., Combaret, V., et al. (2002) Quantification of *MYCN*, *DDX1*, and *NAG* gene copy number in neuroblastoma using a real-time quantitative PCR assay. *Mod. Pathol.* **15**, 159–166.
7. Vandesompele, J., Speleman, F., Van Roy, N., et al. (2001) Multicentre analysis of patterns of DNA gains and losses in 204 neuroblastoma tumors: how many genetic subgroups are there? *Med. Pediatr. Oncol.* **36**, 5–10.
8. Pattyn, F., Speleman, F., De Paepe, A., and Vandesompele, J. (2003) RTPrimerDB: the real-time PCR primer and probe database. *Nucleic Acids Res.* **31**, 122–123.
9. Zuker, M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* **31**, 3406–3415.
10. Livak, K. J. and Schmittgen, T. D. (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* **25**, 402–408.
11. Ririe, K. M., Rasmussen, R. P., and Wittwer, C. T. (1997) Product differentiation by analysis of DNA melting curves during the polymerase chain reaction. *Anal. Biochem.* **245**, 154–160.
12. Vandesompele, J., De Preter, K., Pattyn, F., et al. (2002) Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol.* **3**, RESEARCH0034.1–0034.11.