Review Article

# Accurate and objective copy number profiling using real-time quantitative PCR

Barbara D'haene [a], Jo Vandesompele [a,b], Jan Hellemans [a,b,*]

[a] Center for Medical Genetics, Ghent University Hospital, Ghent, Belgium
[b] Biogazelle, Ghent, Belgium

## ARTICLE INFO

## ABSTRACT

Copy number changes are known to be involved in numerous human genetic disorders. In this context, qPCR-based copy number screening may serve as the method of choice for targeted screening of the relevant disease genes and their surrounding regulatory landscapes. qPCR has many advantages over alternative methods, such as its low consumable and instrumentation costs, fast turnaround and assay development time, high sensitivity and open format (independent of a single supplier). In this chapter we provide all relevant information for a successfully implement of qPCR-based copy number analysis. We emphasize the significance of thorough in silico and empirical validation of the primers, the need for a well thought-out experiment design, and the importance of quality controls along the entire workflow. Furthermore, we suggest an appropriate and practical way to calculate copy numbers and to objectively interpret the results.

The provided guidelines will most certainly improve the quality and reliability of your qPCR-based copy number screening.

## 1. Introduction

Copy number changes under the form of deletions and duplications are known to be involved in numerous human genetic disorders. Moreover, each individual's genome embodies several copy number polymorphisms of various sizes which are thought to contribute to normal phenotypic variation and susceptibility to multifactorial disease [1,2]. Hence, it is not surprising that a wide spectrum of laboratory methods has been developed to identify these copy number changes. Well known and widely applied techniques include conventional karyotyping, fluorescent in situ hybridization (FISH) analysis, microarray-based copy number screening, multiplex ligation-dependent probe amplification (MLPA), and quantitative PCR (qPCR) [3,4]. Each method is characterized by particular (dis)advantages and the choice for a given technique largely depends on the application, required resolution, flexibility, workload, and cost. Conventional karyotyping allows the detection of structural variations across the entire genome, but it is limited in resolution (>5–10 Mb). FISH analysis for targeted regions has been used in a routine setting for many years, and requires either metaphase chromosomes (similar to karyotyping) or interphase nuclei (resolution approximately 100 kb). Microarray-based copy number profiling has improved the resolution in the last decade and facilitated the detection of much smaller copy

number changes [4]. The most recent high density targeted arrays even achieve a resolution of a few base pairs. For patients with mental retardation or other complex phenotypes, genome wide copy number profiling using microarrays proves to be the most suitable approach to reveal the underlying molecular defect [5]. In contrast to the research driven race for ever increasing resolution, the majority of diagnostic tests for genetic disorders are restricted to the targeted screening of the relevant disease genes and their surrounding regulatory landscapes. In the latter context, focused copy number screening methods are preferentially used, such as MLPA, targeted microarrays, and qPCR.

With real-time qPCR, the PCR product accumulation is measured in real-time resulting in a sigmoidal amplification curve. Several detection chemistries are available to measure product accumulation, including hydrolysis probes, molecular beacons, dual hybridization probes and double stranded DNA specific binding dyes. There is a relationship between the moment that the fluorescent PCR signal increases above the background and the initial amount of input DNA; larger amounts of input material will result in lower quantification cycle ($C_q$) values. The $C_q$ value represents the fractional PCR cycle that is characteristic for the amplification curve (e.g. where increase in fluorescence is maximum) or at which the fluorescence crosses a certain threshold. qPCR has many advantages over alternative methods, such as its low consumable and instrumentation costs, fast turnaround and assay development time, high sensitivity and open format (independent of a single supplier). To date, qPCR is the golden standard for gene expression analysis. For copy number determination, qPCR has been less

* Corresponding author. Address: Center for Medical Genetics, Ghent University Hospital, De Pintelaan 185, B-9000 Ghent, Belgium. Fax: +32 9 3326549.
E-mail address: Jan.Hellemans@UGent.be (J. Hellemans).

frequently used, but recent developments hold the promise of taking this application to the next level.

In this chapter we provide all relevant information required to use qPCR for copy number analysis.

## 2. Description of the method

### 2.1. Selection of regions

The overall accuracy of an answer to a specific research or diagnostic question largely depends on the number of qPCR assays, their position relative to the disease locus, and the spatial interval between subsequent amplicons. First, the genes or intergenic regions of interest are selected. The number of selected genes usually affects the number of assays included per gene. When a large series of genes needs to be screened, the number of assays per gene is often restricted to one or two due to practical and financial constraints. If a copy number variant (CNV) in a given gene has already been associated to a specific disorder or phenotype, then such a gene may need to be screened in much more detail to ensure the detection of small deletions. Ultimately, one often wants to screen for CNVs in every single exon, taking into account the differentially spliced exons. The latter would require the development of at least one assay for each exon. Increasing the number of amplicons will increase both the resolution and the cost per sample. Hence, there is a tradeoff between cost per sample and the CNV detection ratio. In conclusion, the number of amplicons and the spacing between successive amplicons largely depends on the specific research or diagnostic question. In addition, it may be important to consider the screening cost per sample.

### 2.2. Assay design and validation

#### 2.2.1. Primer design and in silico quality control

Dedicated design of the qPCR primers and extensive *in silico* and empirical validation are the key for success.

The proper selection of the region to be quantified is obviously the starting point for primer design. The DNA sequence for a gene or for a genomic region can easily be retrieved from the UCSC genome browser (http://genome.ucsc.edu). The genomic position of a region or the name of a gene can be entered in the 'Human Genome Browser Gateway'. In case of a gene name, the browser allows you to select a specific RefSeq sequence. In the actual Genome Browser window you can customize the annotation tracks that are of interest, in this case the 'RefSeq Genes' and the latest 'SNP' track. Next, the DNA sequence can be retrieved by clicking on the DNA link on the top blue menu bar in the Genome Browser. This tool allows custom configuration of the DNA display, which can be useful to highlight exonic regions or mask repeats. Protein domains or other sequences that can be expected to have similarities with other genomic regions are to be avoided to reduce the chance of nonspecific amplification.

Primer pairs for the region of interest should be designed according to stringent parameters to ensure successful assays and convenient experiment design. Practical primer design tools for this application are the PrimerQuest software (Integrated DNA Technologies, http://eu.idtdna.com/scitools/applications/primerquest/default.aspx) and Primer3Plus (http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi). Primer design parameters:

- Primers length: 9 bp/20 bp/30 bp (minimum/optimal/maximum).
- Melting temperature ($T_m$) primers: 58 °C/59 °C/60 °C (minimum/optimal/maximum).
- Maximum $T_m$ difference between primer pairs: 2 °C.
- GC content primers: 30%/50%/80% (minimum/optimal/maximum).

- Amplicon length: 80–150 bp.
- The five nucleotides at the 3′-end should have no more than two G or C bases. This can be set by adjusting the max 3′ stability parameter.
- Practice has shown that the optimal maximum 3′ stability value is 3, but higher values may generate acceptable primers as well.
- Avoid runs of four or more identical nucleotides (especially G bases).

Subsequently, the generated primers should be subjected to *in silico* validation to avoid secondary structures, single nucleotide polymorphisms (SNPs) and copy number polymorphisms at the annealing sites [6,7]. Secondary structures may exert a profound influence on the PCR efficiency and hence adversely affect accurate copy number analysis. The absence of secondary structures in the region in which the primers anneal can be verified using MFOLD (http://frontend.bioinfo.rpi.edu/applications/mfold/cgi-bin/dna-form1.cgi), using the theoretical melting temperature of primers and the appropriate Na+ and $Mg^{2+}$ concentrations (depending on the used SYBR Green kit—an average $Mg^{2+}$ concentration of 1.5 mM and 50 mM for Na+ can be used). The formation of secondary structures is favoured in case of negative free energies ($\Delta G$, kcal/mole). Therefore it is recommended to avoid amplicons with $-\Delta G$ values (Fig. 1).

SNPs and copy number polymorphisms can be excluded at the annealing sites using the corresponding tracks from the UCSC browser. The most practical way to do this is by using the *in silico* PCR tool to open the Genome Browser at the position of the amplicon. Subsequently, the presence of polymorphisms can be verified by retrieving the DNA sequence (as described above) and highlighting these unwanted features. Noteworthy, only the underlying features in the displayed annotation tracks window can be marked. Therefore, it is important to display the respective 'variation and repeat' tracks. For regions encompassing known genes, it is also recommended to check for known SNPs registered in the Ensembl Genome Browser by selecting 'sequence' within the gene-based display list (http://www.ensembl.org). Subsequently, it is possible to configure the given 'marked up gene sequences' and highlight known variations within this sequence.

The BLAST program from the NCBI browser (http://www.ncbi.nlm.nih.gov/BLAST/) is used for *in silico* specificity analysis. A primer pair may be considered as specific if the following requirements are fulfilled:

- An expect value close to zero.
- An identities value of 100% for both the forward and reverse primer.
- Primers should be located on complementary strands.
- An amplicon length between 80 and 150 bp.
- Primers should only match via BLAST analysis at the sequence of interest.

RTPrimerDB (http://www.rtprimerdb.org) is an integrative publicly available webtool for the storage, retrieval and analysis of qPCR primer and probe information, and also enables rapid design of primers for gene expression analysis [7,8]. The RTprimerDB assay quality control pipeline integrates several programs to assess the specificity of the designed primers and to detect above mentioned features that negatively affect the amplification efficiency. While RTPrimerDB is currently restricted to the design and evaluation of primers for gene expression analysis of known RefSeq genes, future versions will allow design and validation of copy number assays (F. Pattyn, personal communication).

Primer pairs that meet the above criteria can be synthesized in combination with standard desalting conditions. Upon arrival, it is recommended to dissolve the primers in nuclease-free water to a
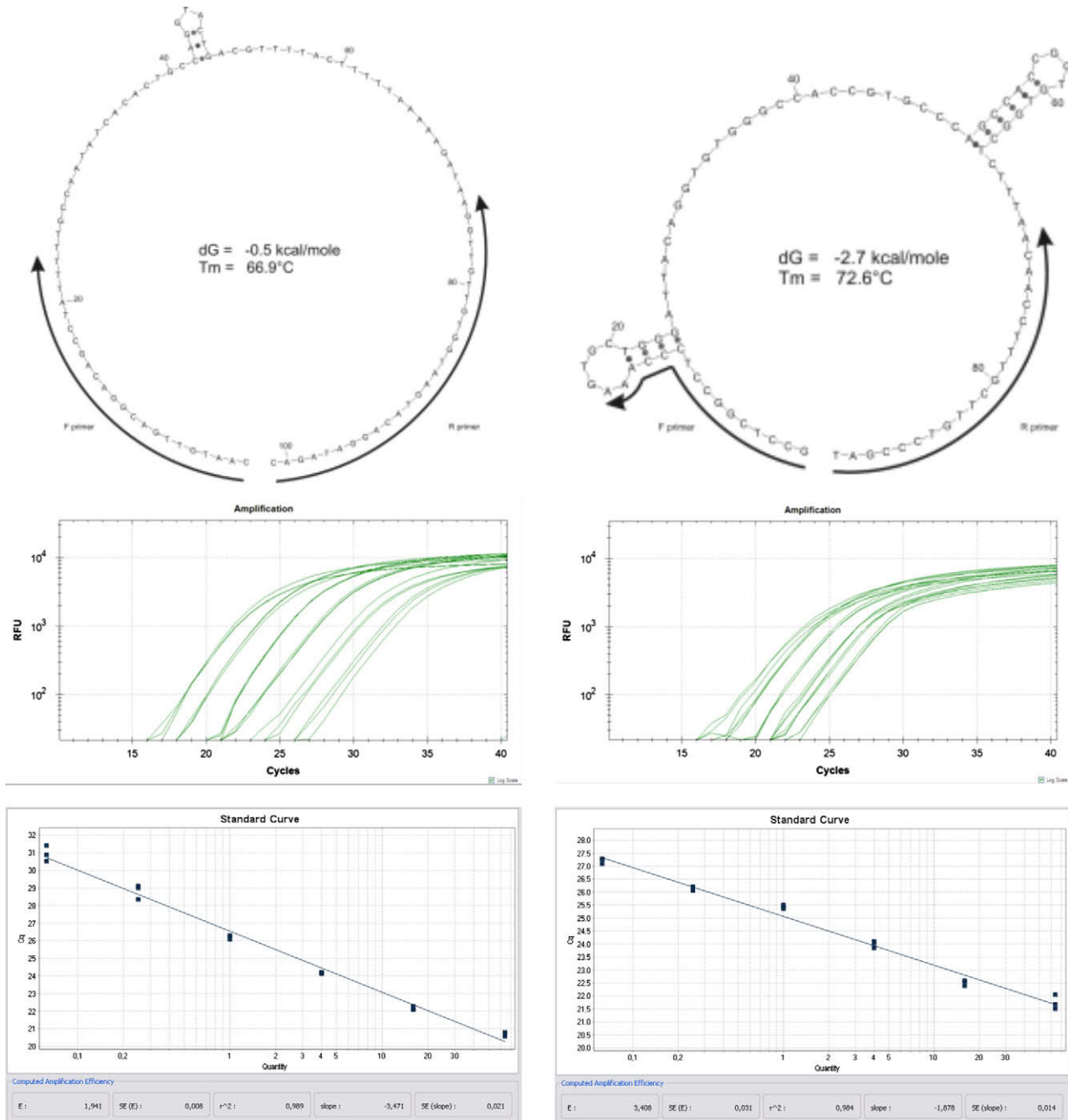
**Fig. 1.** Secondary structure analysis. Secondary structure of an amplicon without (top left) or with secondary structures overlapping the primer annealing sites (top right). The amplification plot on the left reveals equal spacing between the different dilution points at the expected distance of 2 cycles for a 4-fold dilution series. The corresponding amplification plot taken from qbase$^{PLUS}$ shows a good linearity ($r^2 > 0.98$) across the entire concentration range. The $E$ value (base of exponential amplification function) of 1.941 (corresponding to an amplification efficiency of 94.1%) falls within the specifications (1.9–2.1) and has a small error on the estimated efficiency (0.8%). The assay with interfering secondary structures on the right results in an abnormal close spacing between the amplification plots of the different dilution points and an impossible amplification efficiency of 240%.

concentration of 250 μM and to store them at −20 °C. To avoid frequent thaw-freeze cycles, smaller volumes of working solution (5 μM) should be prepared.

### 2.2.2. Empirical validation of primers

After thorough *in silico* quality control, an extensive empirical validation of the primer pairs is required. First, amplification efficiencies are calculated based upon the generation of standard curves using genomic DNA (gDNA) dilution series. Subsequently, melting curve analysis, agarose gel electrophoresis or microchip electrophoresis, and sequencing analysis can be used to check the specificity of the PCR reactions.

#### 2.2.2.1. Preparation of gDNA dilution series for qPCR assay evaluation. Materials

- Human genomic DNA (Roche #1691112, 100 μg, 0.2 μg/μl).
- tRNA from brewer's yeast (Roche #10 109 517 001, 100 mg, lyophilized).
- Nuclease-free water (Sigma #W-4502).

Methods

- Prepare 50× and 1× carrier (tRNA) solutions.
  - ○ Dilute 9 µg carrier in nuclease-free water to obtain a final volume of 36 µl 50× carrier solution (250 ng/µl) and vortex thoroughly
  - ○ Dilute 29 µl 50X solution with 1421 µl nuclease-free water to obtain 1× carrier solution (5 ng/µl) and vortex thoroughly.
- Dilute 50 µl 'Human Genomic DNA' (10 µg) with 6.25 µl 50× carrier and 256.25 µl nuclease-free water and vortex thoroughly (point 1).
- Dilute 78 µl of point 1 solution with 234 µl 1× carrier and vortex thoroughly (point 2).
- Repeat the previous step four times to obtain the dilution points 3–6.
- Add 240 µl of 1× carrier to a tube marked as negative control 1.
- Use a tube with 240 µl $H_2O$ as negative control 2.
- Aliquot all solutions in six equal parts (approximately 40 µl each) and store at −20 °C.

### 2.2.2.2. qPCR analysis using the gDNA dilution series and SYBR Green I. Materials

- Regular pipet, repetition pipet, and multichannel pipet (all regularly calibrated).
- White microtiter plates (appear to result in more precise data than clear plates—own unpublished data).
- *Real-time PCR instrument.* 384-Well systems are to be preferred because they lower both the consumable cost and turn around time (we have good experience with CFX384, LC480 and 7900HT).
- SYBR Green I reagents (many good kits on the market—evaluate for yourself Cq, linearity, efficiency, specificity).
- Nuclease-free water (Sigma).

Methods

- PCR reaction mix for 384-well systems (double the volumes if the screening is to be performed in a 96-well or rotor system).
- Prepare a master mix containing the following components per 7.5 µl PCR reaction:
  - ○ 3.75 µl 2× master mix.
  - ○ 0.375 µl forward primer (working solutions of 5 µM give most of the time good results).
  - ○ 0.375 µl reverse primer (working solutions of 5 µM give most of the time good results).
  - ○ 1 µl nuclease-free water.
- Distribute master mix in 384-well plate (if possible using a repetition pipet).
- Add 2 µl template for each reaction and perform all reactions in duplicate (7 × 2 reactions).
- Close tubes or seal plate (do not write on seal).
- Briefly spin down the 384-well plate.
- Universal PCR protocol:
  - ○ 10 min 95 °C activation hot-start enzyme
  - ○ 40 cycles of 15 s at 95 °C and 1 min at 60 °C
  - ○ dissociation run from 60 to 95 °C (melting curve analysis)

Cq values are extracted with the qPCR instrument software and subsequently imported into qbase$^{PLUS}$ (http://www.qbaseplus.com) for quality control and generation of the standard curves (Fig. 1). For the standard curves the Cq values are plotted against the log concentration of the gDNA template and a linear trend line is fit to the data. The slope of the trend line is used to deduce the PCR efficiency. An optimal efficiency (100%), reflecting the doubling of the PCR product each cycle, is characterized by a slope of −3.32. In general, assays with amplification efficiencies between 90% and 110% are considered as acceptable. In addition to the slope of the standard curve, its linearity (represented by the correlation coefficient $r^2$) and Y-intercept contain valuable quality control information as well. High-quality assays should be linear across the entire dilution series (e.g. $r^2$ close to 1), and the gDNA concentrations of the samples to be tested should fall within this linear range. The Y-intercept should be similar for all CNV assays (e.g. typically within a two cycle interval for our assays).

The inclusion of a dissociation run at the end of the PCR program allows the generation of melting curves. Melting curves reflect the dissociation of the double stranded PCR products and hence can be used to assess the specificity of the PCR reaction. A specific PCR reaction is characterized by a single sharp peak; primer–dimers may generate an additional broader and smaller peak at a lower temperature (Fig. 2).

### 2.3. Experiment design

Experiment design is an indispensible (but often overlooked) step in the workflow of accurate real-time PCR based quantification. The actual design depends on the available qPCR instrument (number of reactions per run), the number of samples (including controls), assays (including references), PCR replicates, and pipetting strategy (manual vs. robotic).

Reference assays are included in the screening to accurately measure and correct for variations in the total amount of input DNA. These assays must amplify a piece of genomic DNA that is known not to be affected (i.e. not registered as a known copy number polymorphism or disease locus). Most autosomal inherited genes with an essential function, not related to the studied phenotype, can be used as a reference sequence. For gene expression assays, the use of multiple validated reference assays is generally considered to be the most reliable way to normalize gene expression levels [9]. Fortunately, the normalization for DNA copy number results is more straightforward because DNA copy numbers vary considerably less than gene expression levels. Therefore, the search for good gDNA reference assays is easier and does not involve an extensive geNorm analysis [9]. We suggest the use of two reference assays for qPCR-based CNV analysis because it provides a good balance between the advantages of including multiple reference assays and keeping the budget under control. Everyone can develop their own set of reference assays, or use assays that have been developed and validated in other laboratories. We have, for example, used assays amplifying ZNF80 and GPR15 genomic DNA as references for normalization of the qPCR data since 2005 (RTprimerDB #1021 and #1022, http://www.rtprimerdb.org) [6].

At least two types of control samples should be included in every qPCR-based copy number analysis. As for all PCR based assays, especially when used in a diagnostic setting, 'no template controls' should be included to detect the presence of contaminating DNA. Specific for qPCR-based copy number analysis, is the inclusion of reference samples with a known copy number. During calculations and result interpretation, these control samples will be used as a reference point (or calibrator) for the determination of the copy numbers. The inclusion of multiple reference samples will result in more accurate results (Fig. 3). One could use two samples with a normal copy number, or one with a normal and the other with a known CNV. The latter CNV sample with known deletion or duplication can both serve as reference point and as positive control for the detection of CNVs (see below).

The use of technical PCR replicates has a number of benefits: (i) it allows for quality control on the precision of the obtained qPCR data; (ii) provides better accuracy; and (iii) allows the generation of reliable results even when an individual qPCR reaction failed. However, it should be noted that the cost for the CNV screening in-
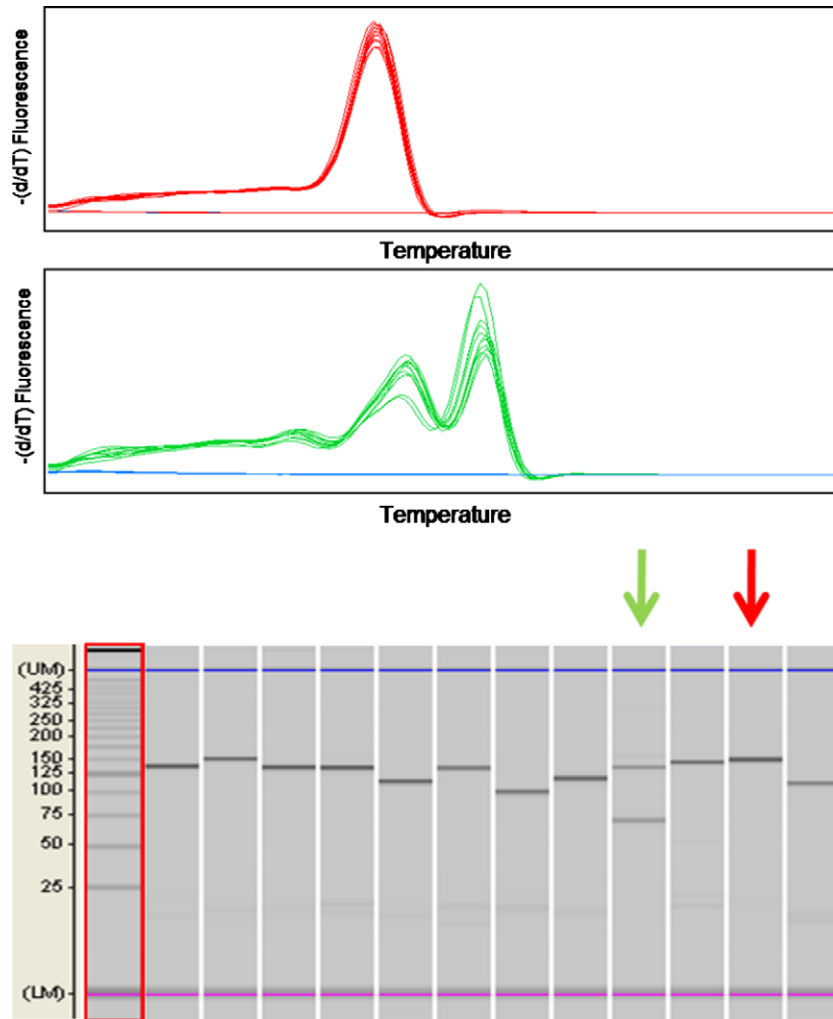
**Fig. 2.** PCR specificity assessment. In a melt curve analysis (top part) the change in fluorescence intensity is plotted in function of temperature. Specific qPCR reactions (red curves) show a single peak whereas multiple peaks can be seen for nonspecific reactions (green curves). The nonspecificity of PCR reactions can also be detected by separating PCR products according to their size on an agarose gel or with microfluidic approaches (bottom part).
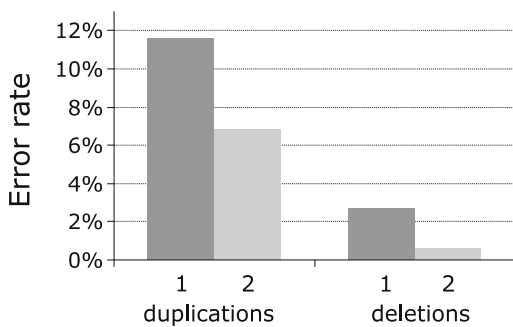


**Fig. 3.** The inclusion of a second reference sample for copy number rescaling has a profound effect on the number of false positive copy number variations. The percentage of false positive duplications or deletions drops from 11.6% to 6.8% (~40% reduction) and from 2.7% to 0.6% (~80% reduction), respectively (results from a single experiment on 40 samples).

creases in proportion to the number of PCR replicates. Strict guidelines on the optimal number of PCR replicates cannot be given since it depends heavily on the quality of the assay, the qPCR instrument, and the Cq determination method, as well as on the pipetting skills of the person performing the qPCR-based CNV screening. Given that all these parameters are of the highest quality, two PCR replicates may be sufficient (see Section 2.5).

Based on this information, the number of reactions for a CNV study can be calculated. For example:

| | |
|---|---|
| *Samples* | 12 |
| Samples of interest | 9 |
| No template control | 1 |
| Reference samples (for copy number rescaling) | 2 |
| *Assays* | 16 |
| Assays covering your region of interest | 14 |
| Reference assays (for normalization) | 2 |
| *Replicates* | 2 |
| Total number of wells | 384 |

This example fits perfectly into a single 384-well plate or in four 96-well plates (Fig. 4). If reactions need to be spread across multiple runs, one needs to use the sample maximization approach to reduce technical variation [10]. This approach dictates that different assays can be spread across as much runs as needed, as long as all samples for a given assay are measured within the same run.

The analysis cost per sample can be reduced by altering the ratio of samples of interest over the total number of samples. For only one sample of interest this ratio is ¼, indicating that ¾ of the analysis cost is due to control samples. For three samples of interest, this ratio improves to ½, giving a 2-fold increase in cost
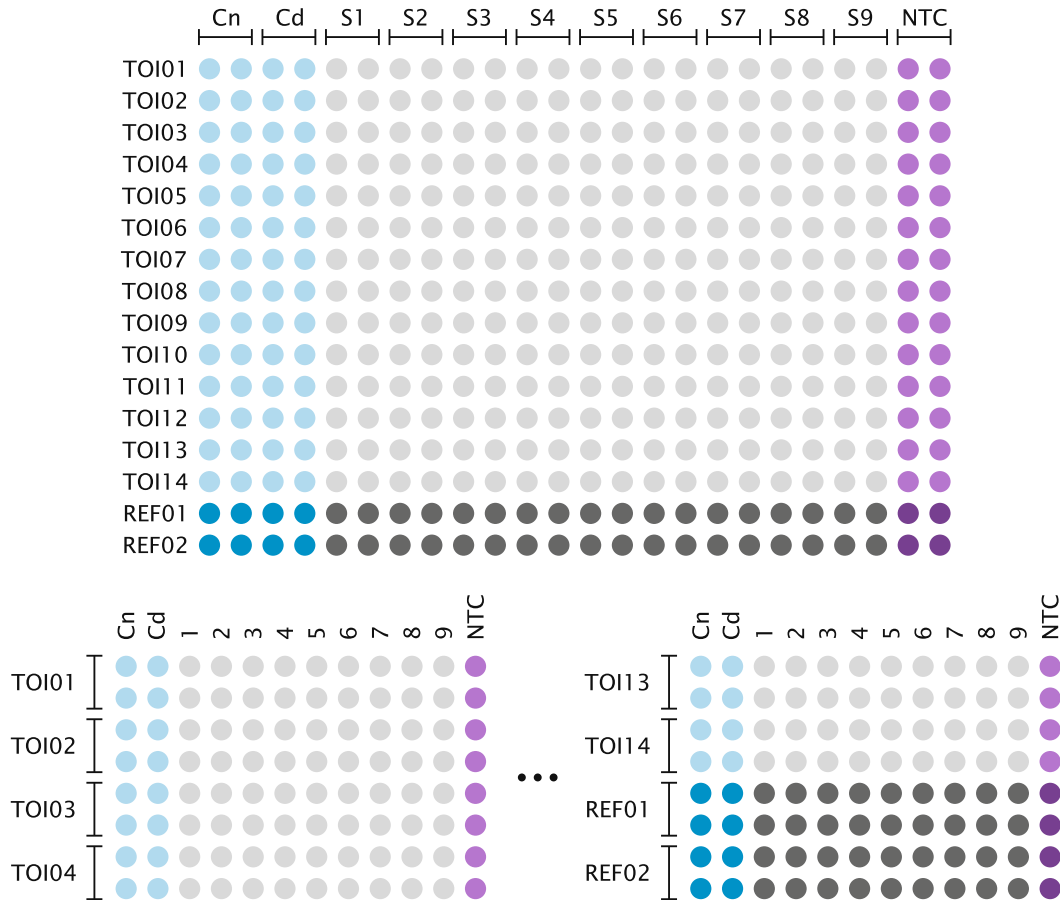
**Fig. 4.** Run lay-out. Schematic representation of the run lay-out for an example screening in a 384-well plate (top) or four 96-well plates (bottom). The sample maximization approach is preferably used in 96-well plates to minimize technical variation between runs [10]. TOI, target of interest; REF, reference target; Cn, normal control; Cd, deletion control; S, unknown sample; NTC, no template control.

efficiency. As the number of samples increases the overhead cost of control samples decreases and becomes marginally for large studies with e.g. more than 100 samples.

### 2.4. qPCR reactions and measurements

The PCR conditions and the subsequent measurements for the actual copy number profiling are the same as for the dilution series (see above). Genomic DNA purified from EDTA blood samples usually serves as an optimal template for qPCR-based copy number profiling. In our hands, DNA from heparin blood samples is of suboptimal quality (see Supplementary Fig. 1), most likely due to the inhibiting nature of heparin.

Cq values are extracted with the qPCR instrument software, exported, and subsequently imported into qbase$^{PLUS}$. The assessment of melting curves allows a rapid control of the specificity. Further control of the specificity using electrophoresis or sequencing is no longer needed.

### 2.5. Copy number calculations

Relative quantities can best be calculated using the universal qBase quantification model that allows for PCR efficiency correction, multiple reference assay normalization, proper error propagation and—if needed—inter-run calibration (http://www.qbaseplus.com) [10]. Several types of quality control are integrated to ensure trustworthy results.

### 2.5.1. Data analysis and quality control procedure in qbase$^{PLUS}$

- Create a new experiment.
- Import run data.
- Annotate the run with sample and assay names (if it has not yet been done in the qPCR instrument software).
- Select reference genes to be used.
- Correct for possible differences in assay PCR efficiency using either a standard curve or a previously determined efficiency.
- Perform replicate quality control. The difference in quantification cycle between the replicate with the highest and lowest Cq value (ΔCq) should be below 0.5 and preferably below 0.3 for all replicates. Replicates that fail this quality control should be inspected carefully. Bad data points can be excluded from data analysis if they can be identified unequivocally, e.g. outlying Cq value (only possible if more than two replicates are included), abnormally high Cq value indicating pipetting error or failed PCR reaction, or abnormal melting curve. Leaving low quality replicates in your dataset may be acceptable, as this will result in a larger error bar on the results.
- Perform a no template quality control. Ideally, the no template control (NTC) should have no Cq value. Amplification signals in the NTC sample indicate contamination issues or primer dimer problems. These problems can be ignored as long as the difference in Cq value between the NTC and the sample with the highest Cq value is large enough. For example, a Cq value difference of five cycles corresponds to a 32-fold difference, indicating that approximately 3% of the signal in your samples is caused by these unwanted signals, which is well below the mea-

surement error and thus perfectly acceptable. Smaller differences in Cq value between the NTC and the other samples are best treated with care. If needed, the entire assay should be repeated.

- Perform reference assay quality control. The geNorm *M* value should be below 0.2 and the coefficient of variation (CV) on the normalized relative quantities (NRQ) for the reference assays should be below 10%. Higher values may indicate problems with the qPCR reactions for the reference assays or copy number polymorphisms in the reference assays that interfere with proper normalization.
- Rescale to a reference sample; most often this is a sample with a normal copy number for the locus under investigation. If a deletion control is selected as a reference the results will represent actual copy numbers (e.g. 2 for a normal diploid locus) rather than relative quantities (e.g. 100% of the quantity found in the reference sample).
- Inspect results visually and export normalized relative quantities for further processing in a spreadsheet (Fig. 5).

### 2.5.2. Calculation of the normal variation in a spreadsheet

- To be able to calculate the assay specific standard deviation, it is recommended to test a sufficiently large (at least 24) normal control series (diploid copy number, CN = 2).
- Rescale the normalized relative quantities to copy numbers (see above and in 2.5.3).
- Log transform copy numbers.
- Use the standard deviation to determine the interval in which 95% of the normal results are expected (NORMINV function in Microsoft Excel and OpenOffice.org Calc). Assuming a similar deviation for samples with a deletion or duplication, 95% intervals can be calculated for copy number 1 or 3 as well.

- Take the anti-log of the obtained intervals for interpretation. Assays for which the 95% intervals for the different copy numbers do not overlap are of sufficient quality to be used in a copy number screening. The intervals for these assays should not overlap the theoretical boundaries between copy number 1, 2 and 3: 1.414 (geometric mean of 1 and 2) and 2.449 (geometric mean of 2 and 3). Some assays may be of sufficient quality to detect deletions but have too much variation to systematically and reliably distinguish normal samples from samples with a duplication.

### 2.5.3. Objective interpretation of results in a spreadsheet (Fig. 6)

- *Rescale to reference samples.* Rescaling has to be performed for each assay individually by dividing the NRQ values (and their error) with an assay specific rescaling factor (RF), defined as the geometric mean of the diploid copy number corrected NRQ value (formula 1; $NRQ_{ij}$, normalized relative quantity from qbase$^{PLUS}$ for sample $i$ and assay $j$; $CN_{ij}$, diploid copy number of the locus measured by assay $j$ in sample $i$). When using only a single reference sample, general formula 1 is simplified to formula 2; when using a normal and deletion control, the rescaling factor is given in formula 3.

$$RF_j = \prod_{i=1}^{n} \frac{NRQ_{ij}}{CN_{ij}} \tag{1}$$

$$RF_j = \frac{NRQ_{norm,j}}{2} \tag{2}$$

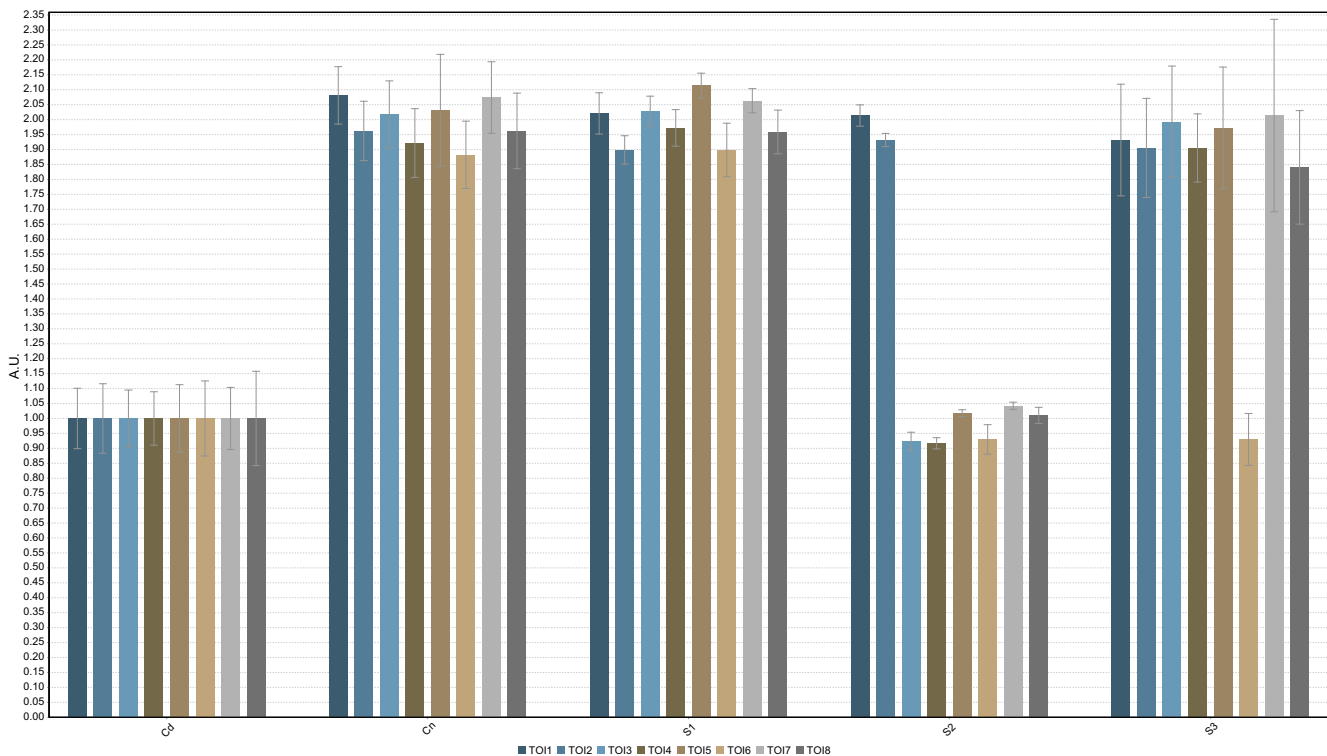$$RF_j = \sqrt{\frac{NRQ_{norm,j}}{2} \cdot \frac{NRQ_{del,j}}{1}} \tag{3}$$

**Fig. 5.** Copy number analysis in qbase$^{PLUS}$. Multi-target bar chart showing the copy numbers in two reference samples (Cn for the normal control and Cd for the deletion control) and three unknown samples for eight different assays after rescaling to Cd. S1 has a normal copy number for all eight assays, whereas S2 and S3 show deletions for 6/8 and 1/8 assays, respectively.

- Determine the most likely copy number for each unknown sample. Results between 1.414 and 2.449 most likely represent a normal copy number of 2. Anything below or above these thresholds is most likely a deletion (CN = 1) or a duplication (CN = 3), respectively.
- For each sample-assay combination, calculate Z-scores on log transformed copy numbers for the most likely interpretation of this copy number (formula 4; $Z_{ij}$ = Z-score for sample $i$ and assay $j$, $CN_{ij}$ = copy number for for sample $i$ and assay $j$, stdev$_j$ = standard deviation of log transformed copy numbers for assay $j$ in normal samples, see above).

$$Z_{ij} = \frac{\log(CN_{ij,calculated}) - \log(CN_{ij,interpretation})}{stdev_j} \qquad (4)$$

- Plot the calculated copy numbers with their error and the Z-scores for the series of assays in two bar charts. Color coding can be applied to these charts to facilitate interpretation of results. For example: red colored bars for assays with an interpreted copy number of 1, grey for normal copy numbers and blue for assays reveling a duplication. Abnormal results (calculated copy numbers outside the predetermined copy number intervals) can be highlighted based on their Z-score (e.g. |Z|>2) to allow objective acceptance or rejection of data.
- Verify that the reference samples show the expected copy number across all assays.

## 3. Concluding remarks

Real-time quantitative PCR is a perfectly suited method for the detection of copy number variations in targeted regions because of its low screening cost and fast turnaround time. Adhering to the following general guidelines increases the quality and reliability of the determined copy numbers. First of all, take care of the preparations: assay design and validation, as well as experiment design. Secondly, perform as much quality controls as possible along the entire workflow: QC on the assay performance and variability, analysis of PCR replicates and evaluation of positive and negative controls. Thirdly, more is better! The inclusion of PCR replicates and more than one reference assay and reference sample improve
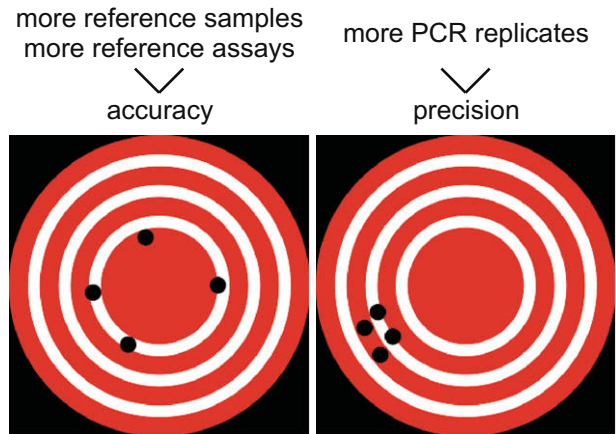


Fig. 7. Improving accuracy and precision. A more accurate determination of the normalization and rescaling factor by using multiple reference assays and reference samples increases the final accuracy of the calculated copy numbers. PCR replicates (duplicates or triplicates) mainly enhance the precision of your results.

both the accuracy and precision of the calculated copy numbers (Fig. 7). Finally, try to make the interpretation of copy numbers as objective as possible, for example by calculating Z-scores for all your results.

Two recent initiatives may help you with the exchange and publication of your qPCR-based CNV results. RDML has been designed as a universal format for the exchange of qPCR data, along with relevant information such as reaction conditions, sample names and assay details [7]. RDML files can be generated on line at the RDML website (http://www.rdml.org) or in RDML supporting software such as qbase$^{PLUS}$. The minimal information about qPCR experiment (MIQE) guidelines contain a checklist of all points to pay attention to when doing real-time quantitative PCR experiments [11]. Make sure to check all the relevant points if you want to publish your results, and for all other experiments that should meet minimal standards.
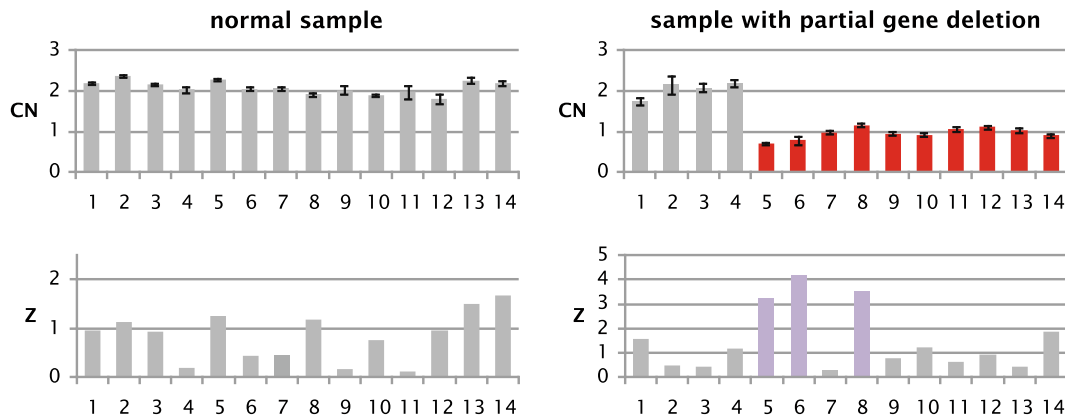
Fig. 6. Objective copy number interpretation. Visualization of copy numbers in a custom bar chart generated by a spreadsheet. Error bars are added to the CN bar chart to allow interpretation of the assay's precision, and a Z-score bar chart is included to provide information about the accuracy of the obtained copy numbers. Color coding is applied to the copy numbers (CN = 1 red, CN = 2 grey, CN = 3 blue) and to the Z-scores (Z < 2 grey, Z > 2 purple) to facilitate interpretation. The results for the sample on the left represent a flawless screening with low technical variability (small error bars) and good accuracy (grey Z-score bars) for a sample with a normal copy number for all assays evaluated (grey CN bars). The results on the right are derived from a sample with a partial gene deletion (red CN bars for assays 5–14). Not all assays give results within the expected CN range (purple Z-score bars). Inaccurate measurements should be repeated unless the surrounding assays provide sufficient information to confirm a deletion or duplication.

## Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ymeth.2009.12.007.

## References

[1] R. Redon, S. Ishikawa, K.R. Fitch, L. Feuk, G.H. Perry, T.D. Andrews, H. Fiegler, M.H. Shapero, A.R. Carson, W. Chen, E.K. Cho, S. Dallaire, J.L. Freeman, J.R. Gonzalez, M. Gratacos, J. Huang, D. Kalaitzopoulos, D. Komura, J.R. MacDonald, C.R. Marshall, R. Mei, L. Montgomery, K. Nishimura, K. Okamura, F. Shen, M.J. Somerville, J. Tchinda, A. Valsesia, C. Woodwark, F. Yang, J. Zhang, T. Zerjal, L. Armengol, D.F. Conrad, X. Estivill, C. Tyler-Smith, N.P. Carter, H. Aburatani, C. Lee, K.W. Jones, S.W. Scherer, M.E. Hurles, Nature 444 (2006) 444–454.
[2] J.R. Lupski, P. Stankiewicz, PLoS Genet. 1 (2005) e49.
[3] J.P. Schouten, C.J. McElgunn, R. Waaijer, D. Zwijnenburg, F. Diepvens, G. Pals, Nucleic Acids Res. 30 (2002) e57.
[4] N.P. Carter, Nat. Genet. 39 (2007) S16–S21.
[5] A.M. Slavotinek, Hum. Genet. 124 (2008) 1–17.
[6] J. Hoebeeck, R. van der Luijt, B. Poppe, E. De Smet, N. Yigit, K. Claes, R. Zewald, G.J. de Jong, A. De Paepe, F. Speleman, J. Vandesompele, Lab. Invest. 85 (2005) 24–33.
[7] S. Lefever, J. Vandesompele, F. Speleman, F. Pattyn, Nucleic Acids Res. 37 (2009) D942–D945.
[8] F. Pattyn, P. Robbrecht, A. De Paepe, F. Speleman, J. Vandesompele, Nucleic Acids Res. 34 (2006) D684–D688.
[9] J. Vandesompele, K. De Preter, F. Pattyn, B. Poppe, N. Van Roy, A. De Paepe, F. Speleman, Genome Biol 3 (2002) RESEARCH0034.
[10] J. Hellemans, G. Mortier, A. De Paepe, F. Speleman, J. Vandesompele, Genome Biol. 8 (2007) R19.
[11] S.A. Bustin, V. Benes, J.A. Garson, J. Hellemans, J. Huggett, M. Kubista, R. Mueller, T. Nolan, M.W. Pfaffl, G.L. Shipley, J. Vandesompele, C.T. Wittwer, Clin. Chem. 55 (2009) 611–622.