Methods 50 (2010) 271–276

Contents lists available at ScienceDirect

Methods

journal homepage: www.elsevier.com/locate/ymeth

Taking qPCR to a higher level: Analysis of CNV reveals the power of high throughput qPCR to enhance quantitative resolution

Suzanne Weaver, Simant Dube, Alain Mir, Jian Qin, Gang Sun, Ramesh Ramakrishnan, Robert C. Jones, Kenneth J. Livak *

Fluidigm Corporation, 7000 Shoreline Court, Suite 100, South San Francisco, CA 94080, USA

ARTICLE INFO

Article history:
Accepted 11 January 2010
Available online 15 January 2010

Keywords:
qPCR
Real-time PCR
Digital PCR
Copy number variation
CNV
Microfluidic array
High throughput qPCR

ABSTRACT

This paper assesses the quantitative resolution of qPCR using copy number variation (CNV) as a paradigm. An error model is developed for real-time qPCR data showing how the precision of CNV determination varies with the number of replicates. Using samples with varying numbers of *X* chromosomes, experimental data demonstrates that real-time qPCR can readily distinguish four copes from five copies, which corresponds to a 1.25-fold difference in relative quantity. Digital PCR is considered as an alternative form of qPCR. For digital PCR, an error model is shown that relates the precision of CNV determination to the number of reaction chambers. The quantitative capability of digital PCR is illustrated with an experiment distinguishing four and five copies of the human gene *MRGPRX1*. For either real-time qPCR or digital PCR, practical application of these models to achieve enhanced quantitative resolution requires use of a high throughput PCR platform that can simultaneously perform thousands of reactions. Comparing the two methods, real-time qPCR has the advantage of throughput and digital PCR has the advantage of simplicity in terms of the assumptions made for data analysis.

1. Introduction

The exponential nature of PCR has had a revolutionary impact on the study of biology. PCR amplification simultaneously addresses both the low quantity of a single copy gene in a sample and the difficulty of specifically detecting that single copy sequence in a highly complex background. The advent of real-time qPCR has added true quantitative ability to the power of PCR. Again, the exponential nature of PCR enables accurate quantification over as many as nine orders of magnitude. Nevertheless, the blessings of exponential amplification become a curse when trying to detect small differences in copy number. Optimized PCRs achieve a doubling of template with each cycle. Correspondingly, qPCR instrument manufacturers specify that their instruments will routinely distinguish 2-fold differences in starting copy number. Because of the doubling per cycle inherent in PCR, distinguishing finer differences than 2-fold requires reliable assessment of fractional cycle differences. This helps explain why literature reports on the limit for qPCR sensitivity range from 1.5- to 2-fold [1,2]. This paper will explore how much better than 2-fold discrimination qPCR can achieve.

Copy number variation (CNV) is an attractive application for examining the resolution of relative quantification. Attempts to identify quality metrics for evaluating gene expression measurements are often frustrated by the lack of standards with known relative quantities of specific transcripts. The MicroArray Quality Control (MAQC) project [3,4] established a framework to assess whether different platforms and laboratories obtain the same answer, but it did not provide a standard for the correct answer. In addition to the lack of standards, assessment of quantification is complicated by the variability introduced due to differing RNA quality and by the added complexity associated with the reverse transcriptase step used in most RNA detection methodologies. In contrast, the nature of germline DNA copy number variation results in relative quantification that is in integer ratios. Thus, the comparison of a sample with four copies of a sequence to a sample with five copies generates exactly a 1.25-fold difference in relative quantity. Furthermore, variability due to sample processing is reduced because CNV analysis is performed directly on genomic DNA. Here, we use human copy number variants as models [5,6] to assess the quantitative resolution of qPCR.

High throughput qPCR analysis has benefited from the development of microfluidic platforms that enable thousands of reactions in a single experiment [7–9]. The ability to simultaneously perform many reactions also has direct bearing on improving the quantitative resolution of qPCR. For real-time qPCR, this paper will present data on the relationship between the number of replicates and the number of copies per genome that can be distinguished. At about the same time that Higuchi was using real-time PCR [10], quantification of DNA using limiting dilutions, or digital PCR, was also

* Corresponding author. Fax: +1 650 871 7152.
E-mail address: ken.livak@fluidigm.com (K.J. Livak).

1046-2023/$ - see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.ymeth.2010.01.003

reported [11,12]. In fact, digital PCR can be thought of as endpoint qPCR. High throughput platforms now make it practical to use digital PCR to detect small differences in copy number [13–17]. For digital PCR, the relationship that will be presented is between the number of reaction chambers and the number of copies per genome that can be distinguished. In addition, this paper will compare the quantitative capabilities of real-time qPCR and digital PCR.

2. Description of method

2.1. Measurement of CNV using real-time qPCR

2.1.1. Statistical analysis

A straightforward way for determining CNV using real-time PCR is to use the $2^{-\Delta\Delta C_q}$ method [18–21]. This method uses a target assay (T) for the DNA segment being interrogated for copy number variation and a reference assay (R) for an internal control segment, which is typically a known single copy gene. For the case where target and reference assays are run in separate reaction chambers, the first step in determining relative copy number is to calculate the average C_q values for each sample for the target and reference assays. The standard deviation (σ) of the average value and the number of measurements (n) that determine the mean is used to calculate the Standard Error of the Mean (SEM).

The best estimate for the C_q of the target assay is:

$$\text{average } C_{q,\mathrm{T}} \pm \frac{\sigma_{\mathrm{T}}}{\sqrt{n_{\mathrm{T}}}} = C_{q,\mathrm{T}} \pm \mathrm{SEM}_{\mathrm{T}} \tag{1}$$

The best estimate for the C_q of the reference assay is:

$$\text{average } C_{q,\mathrm{R}} \pm \frac{\sigma_{\mathrm{R}}}{\sqrt{n_{\mathrm{R}}}} = C_{q,\mathrm{R}} \pm \mathrm{SEM}_{\mathrm{R}} \tag{2}$$

The ΔC_q value ($C_{q,\mathrm{T}} - C_{q,\mathrm{R}}$) is a measure of the copy number of the target segment relative to the reference segment. The use of ΔC_q normalizes for differences in input concentration when comparing different samples. The uncertainty in ΔC_q is the square root of the quadratic sum of the uncertainties in the individual C_q values, assuming the uncertainties are independent and random [22]. The uncertainties are independent because each set of C_q values is derived from individual, independent reactions. In addition, we observe no significant systematic bias in the qPCR replicates or platform that would indicate the assumption about being random is invalid.

$$\mathrm{SEM}_{\Delta C_q} = \sqrt{\mathrm{SEM}_{\mathrm{T}}^2 + \mathrm{SEM}_{\mathrm{R}}^2} \tag{3}$$

The next step in determining relative copy number is to calibrate each ΔC_q value to a sample with a known copy number for the target segment. Typically, the copy number of the target segment in the calibrator sample (C) is single copy per haploid genome (two copies per diploid genome). Calibration is done by calculating $\Delta\Delta C_q$

$$\begin{aligned}\Delta\Delta C_q &= (\Delta C_q - \Delta C_{q,c}) \pm \sqrt{\mathrm{SEM}_{\Delta C_q}^2 + \mathrm{SEM}_{\Delta C_{q,c}}^2} \\ &= \Delta\Delta C_q \pm \mathrm{SEM}_{\Delta\Delta C_q}\end{aligned} \tag{4}$$

Assuming the efficiencies of the target and reference assays are similar and close to 1 [18], relative copy number (RCN) is calculated from the $\Delta\Delta C_q$ value using the formula:

$$\mathrm{RCN} = 2^{-\Delta\Delta C_q} \tag{5}$$

For the case where the calibrator has two copies of the target segment per diploid genome:

$$\#\text{ of copies per diploid genome} = 2 * 2^{-\Delta\Delta C_q} \tag{6}$$

The $\mathrm{SEM}_{\Delta\Delta C_q}$ value can be used to calculate the 95% confidence interval for each RCN value. The 95% confidence interval for $\Delta\Delta C_q$ is given by $\pm t * \mathrm{SEM}_{\Delta\Delta C_q}$, where t is the appropriate critical value for a t-distribution (two-tailed test with $p = 0.05$). Thus, the 95% upper and lower bounds for RCN are:

$$\mathrm{RCN}_{\max} = 2^{-\Delta\Delta C_q + t * \mathrm{SEM}_{\Delta\Delta C_q}} \tag{7}$$

$$\mathrm{RCN}_{\min} = 2^{-\Delta\Delta C_q - t * \mathrm{SEM}_{\Delta\Delta C_q}} \tag{8}$$

These equations assume that the $\Delta\Delta C_q$ values are normally distributed. When a relatively large number of replicates are being used to determine $\Delta\Delta C_q$, this seems a reasonable assumption based on the central limit theorem. The critical t value is used to compensate for experiments with fewer replicates.

As shown in Fig. 1A, these equations can be used to generate a model relating the 95% confidence limit range to the number of replicates per assay per sample. A 95% confidence interval (CI) means that there is a 5% chance that the measured value using the designated number of replicates will be outside the 95% CI. This means there is a 2.5% chance that the value will be greater than the maximum value and a 2.5% chance the value will be lower than the minimum value. For example, at the point where the 4-copy CI and 5-copy CI cross (copy number of approx. 4.47), there is a 97.5% chance that the measured value for the 4-copy sample will be less than the crossover value and a 97.5% chance the measured value for the 5-copy sample will be greater than the crossover value. Thus, the probability that both the 4-copy value is less than the crossover value and the 5-copy value is greater than the crossover value is $0.975 \times 0.975 = 0.950625$. Assuming a system σ of 0.16, the crossover points in Fig. 1A indicate that it should be possible to distinguish, with at least 95.1% probability, one copy from two copies with 5 replicates; two copies from three copies with 8 replicates; three copies from four copies with 12 replicates; and four copies from five copies with 18 replicates. At a system σ of 0.25, similar analysis indicates that one copy can be distinguished from two copies with 7 replicates; two copies from three copies with 14 replicates; three copies from four copies with 26 replicates; and four copies from five copies with 40 replicates.

Kubista et al. [23] describe how to use the Power test to estimate the number of replicates needed in order to detect a certain difference in copy number. The analysis that is equivalent to the analysis depicted in Fig. 1A is a one-tailed t-test with 2.5% false positives (significance) and 2.5% false negatives (power). Fig. 1B shows the results of this Power test. This analysis indicates that, with a system σ of 0.16, it should be possible to distinguish one copy from two copies with 4 replicates; two copies from three copies with 6 replicates; three copies from four copies with 11 replicates; and four copies from five copies with 17 replicates. At a system σ of 0.25, the analysis indicates that one copy can be distinguished from two copies with 6 replicates; two copies from three copies with 13 replicates; three copies from four copies with 24 replicates; and four copies from five copies with 39 replicates. These numbers are slightly smaller than the number of replicates indicated by the analysis of Fig. 1A. This is because the Power test is set to satisfy the condition that the measured value of the lower copy number sample will be lower than the measured value of the higher copy number sample, which is slightly less stringent than requiring that the lower copy value is less than the crossover value and the higher copy value is greater than the crossover value. Also, the significance of the analysis of Fig. 1A is slightly higher than 95%.

2.1.2. Example

Human genomic DNA samples were obtained from Coriell that come from cell lines containing one, two, three, four, or five *X* chromosomes. These samples were analyzed using a target assay for the *YY2* gene on the *X* chromosome and a reference assay for the *RPPH1* gene on chromosome 14. As shown in Fig. 2, these assays were checked in order to confirm that their efficiencies are similar

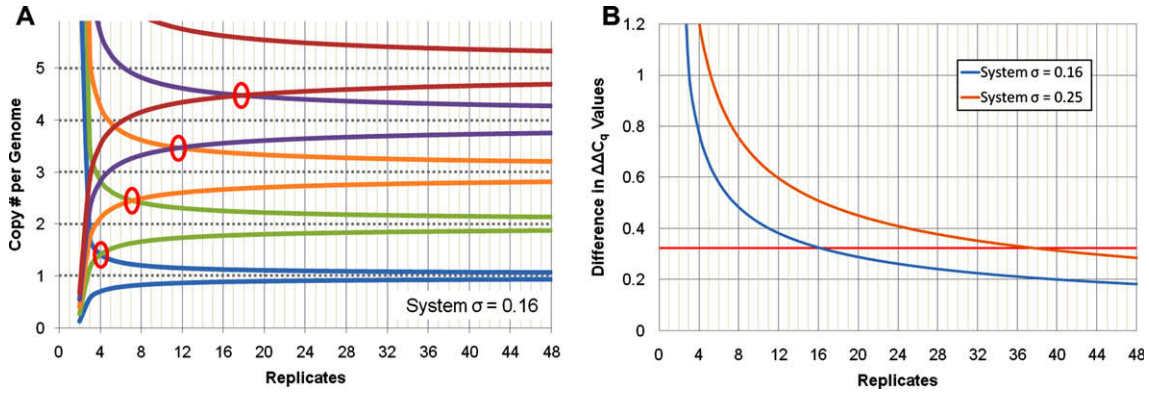

Fig. 1. Error model for real-time qPCR relating 95% confidence interval for copy number determination versus number of replicates. From Eqs. (1)–(4):

$$\mathrm{SEM}_{\Delta\Delta C_q} = \sqrt{\frac{\sigma_{\mathrm{T}}^2}{n_{\mathrm{T}}} + \frac{\sigma_{\mathrm{R}}^2}{n_{\mathrm{R}}} + \frac{\sigma_{\mathrm{T,C}}^2}{n_{\mathrm{T,C}}} + \frac{\sigma_{\mathrm{R,C}}^2}{n_{\mathrm{R,C}}}} \tag{9}$$

If all standard deviations are assumed to be equal to the system standard deviation σ and the same number of replicates (n) are used for the target and reference assays across all samples including the calibrator, then:

$$\mathrm{SEM}_{\Delta\Delta C_q} = \frac{\sqrt{\sigma^2+\sigma^2+\sigma^2+\sigma^2}}{\sqrt{n}} = \frac{\sqrt{4\sigma^2}}{\sqrt{n}} = \frac{2\sigma}{\sqrt{n}} \tag{10}$$

This value for $\mathrm{SEM}_{\Delta\Delta C_q}$ is plugged into Eqs. (7) and (8) and evaluated for a specific value of σ and differing values of n. (A) Plot of $2 \times \mathrm{RCN}_{\min}$ and $2 \times \mathrm{RCN}_{\max}$ versus number of replicates for $\sigma = 0.16$ and $\Delta\Delta C_q = -1$, 0, 0.585, 1, and 1.322 (copy number per genome of 1, 2, 3, 4, and 5 assuming reference sequence is present at 2 copies per genome). (B) Plot of values generated using the 'Exp. Design' tab of GenEx Enterprise software (MultiD Analyses AB, Göteborg, Sweden). The settings used were: 'Number of samples in each group' = 2–48; 'Power (%)' = 97.5; 'Significance' = 95%, 2 tail; 'Type of test' = Unpaired; 'SD Estimated' = Yes (t-test). This is equivalent to a significance of 97.5%, 1 tail, an entry the software does not allow. For system $\sigma = 0.16$, the values entered in 'SD (Group A)' and 'SD (Group B)' were 0.32. This is because, if $\sigma = 0.16$ for C_q values, then the σ for $\Delta\Delta C_q$ values is $\sqrt{(0.16)^2+(0.16)^2+(0.16)^2+(0.16)^2} = 0.32$. For system $\sigma = 0.25$, the values entered in 'SD (Group A)' and 'SD (Group B)' were 0.5. Distinguishing one copy from two copies corresponds to a difference in $\Delta\Delta C_q$ values of 1.0; distinguishing two from three corresponds to a difference of 0.585; distinguishing three from four corresponds to a difference of 0.415; and distinguishing four from five corresponds to a difference of 0.322. The red line indicates a difference value of 0.322, which is the difference required to distinguish four and five copies.

and close to 1. Copy number qPCR analysis was performed in a Fluidigm® 96.96 Dynamic Array integrated fluidic circuit (IFC). In this device, up to 96 samples and 96 assays can be loaded into the matrix array to create 9216 individual reactions. For this analysis of copy number of an *X*-lined gene, each sample was pipetted into 19 Sample Inlets and each assay was pipetted into 24 Assay Inlets for a total of $19 \times 24 = 456$ replicates. The range of standard deviations observed for *YY2* C_q measurements across the five samples was 0.035–0.066, and for *RPPH1* C_q measurements was 0.070–0.096. Table 1 shows the measured copy number for the *YY2* gene relative to the *RPPH1* gene for the five samples. Because of the very large number of replicates, the 95% confidence intervals for these values are very small, demonstrating that qPCR can deliver very precise results.

In order to explore the applicability of the error model in Fig. 1, the experimental data set was used to see how precision of the estimate of copy number varies with the number of replicates. This was done by randomly selecting C_q values from each set of 456 values to create "samples" with a lower number of replicates. The results of one such trial are shown in Fig. 3. In this case, 4 replicates were able to distinguish one copy from two copies and two copies from three copies; 7 replicates were able to distinguish three copies from four copies; and 12 replicates were able to distinguish four copies from five copies. Fewer replicates were required than predicted by the model in Fig. 1A because the system σ is considerably less than 0.16. In order to examine the robustness of the method, the analysis was repeated but this time each set of replicates always included the minimum and maximum value out of all 456 values. The other replicates in each set were again selected randomly. The results of one trial where the minimum and maximum values were deliberately included were that one copy was distinguished from two copies with 6 replicates; two copies from three copies with 7 replicates; three copies from four copies with 16 replicates; and four copies from five copies with 24 replicates. These results indicate that the error model illustrated in Fig. 1 provides a reasonable expectation of the precision that can be obtained with real-time qPCR.

The derivation of the error model focuses on the precision of qPCR, but does not directly address accuracy. The results in Table 1 show that the accuracy of copy number determination for the one-*X*, three-*X*, and four-*X* samples is quite good. For the five-*X* sample, though, the measured copy number is 4.56 versus an expected value of 5. This indicates the existence of an unidentified systematic error that is impairing accuracy for this particular sample. For example, if the five copies of *YY2* in the five-*X* sample are not identical and if the sequence variation in one or more of these copies adversely affects the performance of the specific assay used in this study, then the measured C_q value for *YY2* will be higher than the value expected if all five copies were identical. In this case, these results do clearly distinguish four copies from five copies, but do not accurately report the true degree of separation. Another possibility is that the karyotype of 5 *X* chromosomes is not completely stable in cell culture so that the DNA was extracted from a mixture of cells with 4 and 5 *X* chromosomes. If this were the case, then the results in Table 1 could very well accurately report the copy number in the experimental samples.

2.2. Measurement of CNV using digital PCR ("endpoint qPCR")

2.2.1. Statistical analysis

In digital PCR, a number of target reactions are performed at low template input such that some reaction chambers are positive and some chambers are negative. The absolute concentration of any target sequence (in molecules/μL) can be calculated by count-

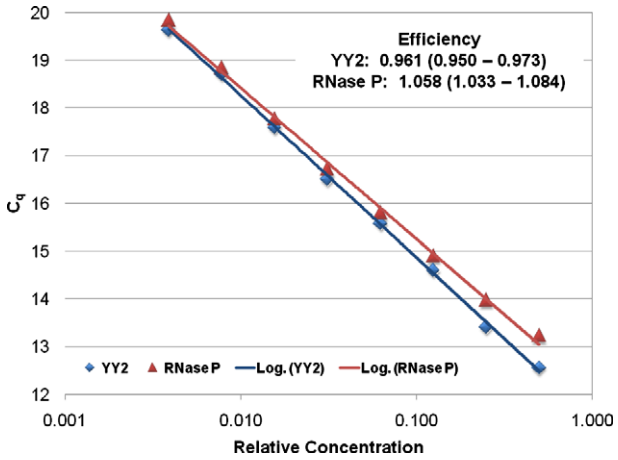

Fig. 2. Standard curves for the *YY2* and *RPPH1* assays. The 20× *YY2* assay consists of primers (CAGTACGAGGATGTGGATGGC and CCTCTTGTGTCTGCAACATAAGC obtained from IDT) at 18 μM each and a hydrolysis probe (FAM-TTCCTGGTCGTGGTCGCCATAGCC-BHQ obtained from Biosearch) at 4 μM. *RPPH1* is the gene encoding the RNA component H1 of RNase P. The 20× *RPPH1* assay was obtained from Applied Biosystems (4316831) and consists of primers at 18 μM each and a FAM-labeled hydrolysis probe at 5 μM. Preamplification was performed in 50 μL reactions containing 4.4 μg genomic DNA, 1× TaqMan® PreAmp Master Mix (Applied Biosystems 4391128) and pooled assays at 0.05× each. Reactions were incubated for 10 min at 95 °C followed by 10 cycles of 15 s at 95 °C/4 min at 60 °C, then diluted 1:5 with H_2O. Eight 2-fold dilutions of this preamplified material were then analyzed separately with each assay in a 96.96 Dynamic Array IFC (Fluidigm). Final reaction conditions consisted of 1× TaqMan Universal PCR Master Mix (Applied Biosystems 4304437), 1× *YY2* or *RPPH1* assay, 1× GE Sample Loading Reagent (Fluidigm 85000735), and 0.1× Assay Loading Reagent (Fluidigm 85000736). The array was analyzed in the BioMark™ real-time PCR system (Fluidigm) using a thermal protocol of 2 min at 50 °C, 30 min at 70 °C, 10 min at 25 °C, 2 min at 50 °C, 10 min at 95 °C, followed by 40 cycles of 15 s at 95 °C/1 min at 60 °C. The C_q values plotted are the average of at least 136 replicates. Efficiency was calculated using the formula

$$\text{efficiency} = 10^{-\frac{1}{\text{slope}}} - 1 \tag{11}$$

The slope and 1-sigma interval were estimated using weighted regression.

ing the number of positive chambers, applying a correction for Poisson distribution, and dividing by the total volume of all the chambers. The relevant parameters for characterizing digital PCR are λ, the average number of target molecules per chamber, and p, the probability that a chamber has at least one target molecule and thus gives a positive PCR. When the number of chambers is large, Sindelka et al. [24] report that λ and p are related by the equation:

$$1 - p = e^{-\lambda} \tag{12}$$

or,

$$\lambda = -\ln(1 - p) \tag{13}$$

The best estimate for p is calculated by dividing the number of positive chambers (H, for hits) by the total number of chambers (C). If C is large enough, the 95% confidence limits for p and λ are given by [17]:

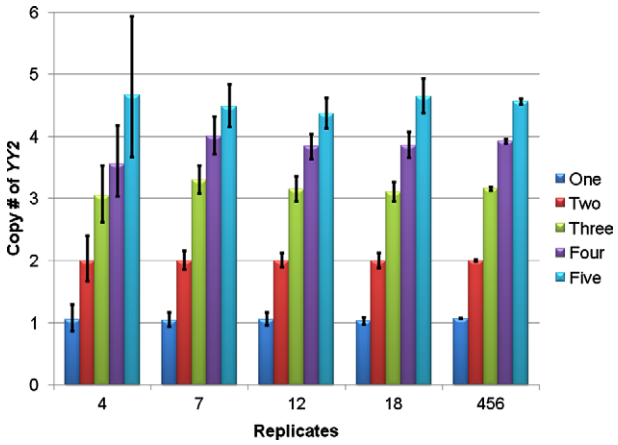

Fig. 3. Copy number analysis using real-time qPCR with variable numbers of replicates. The data used to generate the results in Table 1 (456 replicates for each assay for each sample) were used to create simulated samples that had variable numbers of replicates. For any particular assay and sample, the 456 C_q values were re-ordered using a random number generator. The top n values were assigned to the simulated sample with n replicates. The list of C_q values was re-ordered with a fresh set of random numbers prior to each selection. Copy number was calculated using Eq. (6). Results are shown for samples with one, two, three, four, or five copies of the X chromosome. Error bars show 95% confidence intervals calculated using Eqs. (7) and (8).

$$p_{\min,\max} = p \mp 1.96 * \sqrt{\frac{p(1-p)}{c}} \tag{14}$$

$$\lambda_{\min} = -\ln(1 - p_{\min}) \quad \text{and} \quad \lambda_{\max} = -\ln(1 - p_{\max}) \tag{15}$$

Fig. 4 shows how the 95% confidence interval $\lambda_{\max} - \lambda_{\min}$ varies with different values for λ and p. It can be seen that the least amount of relative error in determining λ occurs when there are approximately 1.5 target molecules per chamber, which corresponds to approximately 80% positive chambers.

Because quantification using digital PCR is based on an absolute measurement of the number of molecules, there is no need to calibrate the results obtained to a sample with known target copy number. Thus, relative copy number (RCN) is determined by comparing the results of the target assay directly to the results of the reference assay for each sample. For the case where the reference sequence is present at two copies per diploid genome:

$$\#\text{ of target copies per diploid genome} = 2 * \frac{\lambda_T}{\lambda_R} \tag{16}$$

The 95% confidence intervals for λ_T and λ_R can be determined using Eqs. (14) and (15). As shown in Fig. 5, these intervals can be used to generate a model relating the 95% confidence limit range for $\frac{\lambda_T}{\lambda_R}$ to the number of reaction chambers. Fig. 5 indicates that one copy can be distinguished from two copies using approximately 200 chambers; two copies from three copies using approximately 400 chambers; three copies from four copies using approximately 800 chambers; four copies from five copies using approximately 1200 chambers, and so on. The final comparison in Fig. 5 indicates that

Table 1
Copy Number of an *X*-Linked Gene Determined by real-time qPCR.[a]

# of *X* chromosomes	One	Two (calibrator)	Three	Four	Five
Measured copy # of *YY2*	1.06	2	3.16	3.92	4.56
95% Confidence interval	1.054–1.076	1.981–2.020	3.132–3.188	3.884–3.960	4.516–4.600

[a] *X* copy variant samples were purchased from Coriell (NA18515, NA18968, NA0623, NA01416, and NA06061 for one, two, three, four, and five copies, respectively). Preamplification and real-time qPCR analysis were performed as described in the legend to Fig. 2 except only one concentration of preamplified DNA was analyzed for each sample. Copy number was calculated using Eq. (6). 95% confidence intervals were calculated using Eqs. (7) and (8) with data from 456 replicates.

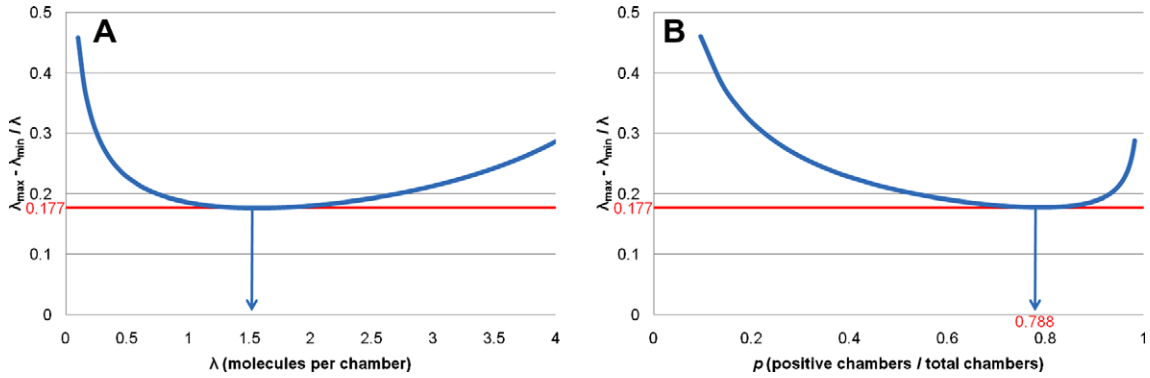

Fig. 4. Relative error in the digital PCR determination of concentration as a function of number of molecules per chamber or fraction of positive chambers. The fraction of positive chambers, p, was evaluated for λ (molecules per chamber) values ranging from 0.1 to 4 using Eq. (12). p_{min} and p_{max} were calculated using Eq. (14) and a value of $C = 770$ chambers. λ_{min} and λ_{max} were calculated using Eq. (15). Relative error was determined by dividing the 95% confidence interval $\lambda_{max} - \lambda_{min}$ by λ. (A) Relative error plotted versus number of molecules per chamber. (B) Relative error plotted versus fraction of positive chambers.

10 copies can be distinguished from 11 copies using approximately 8000 chambers.

2.3. Example

Hosono et al. [6] characterized copy number variation affecting the *MRGPRX1* gene on human chromosome 11. Genomic DNA samples were obtained from Coriell that come from cell lines containing four or five copies of *MRGPRX1*. These samples were analyzed by digital PCR using an assay for *MRGPRX1* with a FAM-labeled probe and an assay for *RPPH1* with a VIC-labeled probe. Differentially labeled probes were used so that the target and reference assays could be run mixed together. Analysis was performed in the Fluidigm 48.770 Digital Array IFC. In this device, the reaction loaded into each of 48 Inlets is subdivided into a panel of 770 chambers. Fig. 6 shows the copy number results when the reactions for each sample were loaded into 24 Inlets. Using the data from all the panels, the measured copy numbers for the four-copy and five-copy samples are 4.12 (95% confidence interval 3.94–4.30) and 5.00 (95% confidence interval 4.84–5.17), respectively. The applicability of the error model in Fig. 5A was examined by ran-

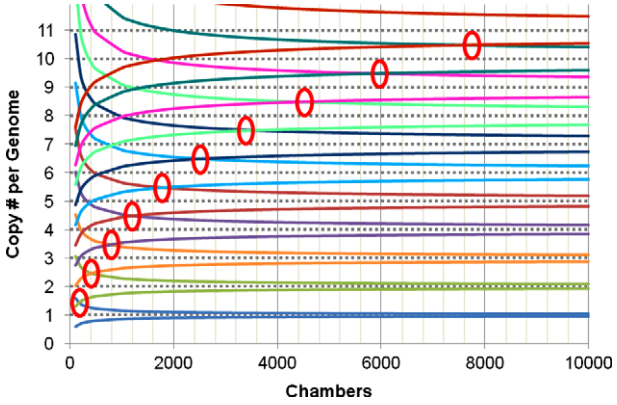

Fig. 5. Error model for digital PCR relating 95% confidence interval for copy number determination versus number of reaction chambers. Let relative copy number $r = \frac{\lambda_T}{\lambda_R}$. As derived in Dube et al. [17], the 95% confidence limits for r are given by:

$$r_{min} = \frac{\lambda_T * \lambda_R - \sqrt{\lambda_T^2 * \lambda_R^2 - ((\lambda_T - \lambda_{T,min})^2 - \lambda_T^2) * ((\lambda_{R,max} - \lambda_R)^2 - \lambda_R^2)}}{\lambda_R^2 - (\lambda_{R,max} - \lambda_R)^2} \quad (17)$$

$$r_{max} = \frac{\lambda_T * \lambda_R + \sqrt{\lambda_T^2 * \lambda_R^2 - ((\lambda_{T,max} - \lambda_T)^2 - \lambda_T^2) * ((\lambda_R - \lambda_{R,min})^2 - \lambda_R^2)}}{\lambda_R^2 - (\lambda_R - \lambda_{R,min})^2} \quad (18)$$

In order to evaluate these expressions, it is assumed that the concentration of genomic DNA is held constant at a level that corresponds to $\lambda_R = 0.6$, i.e., 0.6 copies of the reference sequence per chamber. Plot of $2 \times r_{min}$ and $2 \times r_{max}$ versus number of reaction chambers for $r = 0.5$, 1, 1.5, 2, 2.5, 3. 3.5. 4, 4.5, 5 and 5.5 (copy number per genome of 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 and 11 assuming reference sequence is present at 2 copies per genome).

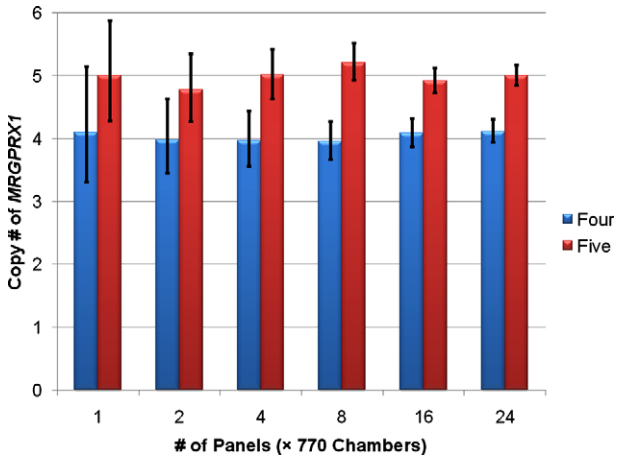

Fig. 6. Copy number of *MRGPRX1* determined by digital PCR. Genomic DNA samples were purchased from Coriell containing either four copies (NA19221) or five copies (NA19205) of *MRGPRX1*. The 20× *MRGPRX1* assay consists of primers (TTAAGCTTCATCAGTATCCCCCA and CAAAGTAGGAAAACATCATCACAGGA obtained from IDT) at 18 μM each and a hydrolysis probe (FAM-ACCATCTCTAAAATCCT-MGB obtained from Applied Biosystems) at 4 μM. The 20× *RPPH1* assay was obtained from Applied Biosystems (4316844) and consists of primers at 18 μM each and a VIC®-labeled hydrolysis probe at 5 μM. Preamplification was performed in 50 μL reactions containing 750 ng genomic DNA, 1× TaqMan PreAmp Master Mix and pooled assays at 0.05× each. Reactions were incubated for 10 min at 95 °C followed by 5 cycles of 15 s at 95 °C/2 min at 60 °C, then diluted 1:50 with H_2O. Digital PCR analysis was performed in a 48.770 Digital Array IFC (Fluidigm) with the four-copy sample loaded in 24 reaction inlets and the five-copy sample loaded in 24 reaction inlets. Final reaction conditions consisted of 1× TaqMan Gene Expression PCR Master Mix (Applied Biosystems 4369016), 1× *MRGPRX1* assay (FAM probe), 1× *RPPH1* assay (VIC probe), and 1× GE Sample Loading Reagent. The array was analyzed in the BioMark real-time PCR system using a thermal protocol of 2 min at 50 °C, 10 min at 95 °C, followed by 40 cycles of 15 s at 95 °C/1 min at 60 °C. Reactions were evaluated using endpoint data. Chambers were called positive if the R_n value was greater than an empirically selected threshold. FAM and VIC results were evaluated independently. $\lambda_{MRGPRX1}$ was calculated using Eq. (13) with p = the total number of FAM hits divided by the total number of chambers. λ_{RPPH1} was calculated using Eq. (13) with p = the total number of VIC hits divided by the total number of chambers. Copy number is $2 * \frac{\lambda_{MRGPRX1}}{\lambda_{RPPH1}}$. Error bars show 95% confidence intervals calculated using Eqs. (17) and (18). Panels were selected randomly to generate the results displayed for 1, 2, 4, 8, and 16 panels.

domly selecting smaller numbers of panels to determine copy numbers. The results of one trial shown in Fig. 6 indicate that the data from 4 panels ($4 \times 770 = 3080$ chambers) are required to distinguish four and five copies. This number is larger than the approximately 1200 chambers predicted by the model in Fig. 5A because the template concentrations used for the experimental data were lower than the concentration used to generate the model. The Fig. 5 model assumes a λ_R of 0.6 reference molecules per chamber. For the experiment presented in Fig. 6, λ_R for the four-copy sample was approximately 0.18 and λ_R for the five-copy sample was approximately 0.37.

3. Concluding remarks

Although CNV was used as the example in this paper, the models for quantitative resolution apply equally well to gene expression studies. Thus, real-time qPCR can be used to routinely distinguish 1.25-fold differences in gene expression as long as one is willing to run 18 (system $\sigma = 0.16$) to 40 (system $\sigma = 0.25$) replicates. In fact, the error model in Fig. 1 predicts that a 1.1-fold difference (10 copies from 11 copies) can be distinguished by running 86 replicates if the system σ is 0.16. System σ includes sample preparation, assay performance, and instrument platform, which were not addressed in this paper. The upstream processes involved in preparing samples and assays for qPCR can result in large variation, so it is important to consider including more than just qPCR technical replicates in any study [25]. If the system σ can be verified to be less than 0.16, then lower numbers of replicates will be required. For digital PCR with the reference template loaded at 0.6 molecules per chamber, the number of chambers required to achieve 1.25-fold discrimination is approximately 1200 chambers, and the number of chambers required to achieve 1.1-fold discrimination is approximately 8000 chambers.

The basic finding of this paper is that the quantitative resolution of qPCR can be enhanced to 1.25-fold, or even 1.1-fold, by running a large number of reactions, either as replicates for real-time qPCR or independent chambers for digital PCR. Although it is possible to achieve this improved quantitative performance using conventional tubes or plates, it is not practical because of the reagent expense involved, the cumbersome workflow engendered by the pipetting required, and the increased chance of introducing variability due to the cumbersome workflow. Thus, high throughput, microfluidic platforms for qPCR are enabling for achieving the utmost in quantitative resolution.

The main advantage of using real-time qPCR over digital PCR is throughput. Consider the case where at least 18 replicates are being run in order to achieve 1.25-fold discrimination. In the Fluidigm 96.96 Dynamic Array IFC, 95 of the 96 Assay Inlets can be used to run 5 different assays with 19 replicates each. Four of these assays could be for separate targets and one assay would be a common reference assay. Thus, a single 96.96 array can be used to analyze 96 samples for 4 targets for a total of $96 \times 4 = 384$ determinations with 95% confidence. For digital PCR, 1.25-fold discrimination can be achieved with approximately 1200 chambers, which corresponds to two panels in a Fluidigm 48.770 Digital Array IFC. Mixing the target and reference assays together, the 48 Reaction Inlets of the 48.770 array can be used to run 24 different samples with two panels per sample. Thus, the throughput of a single 48.770 array is 24 samples for 1 target for a total of 24 determinations.

The advantages of digital PCR over real-time qPCR are the option to run as an endpoint assay and the simplicity of analysis. This simplicity stems from the fact that quantification is based on the counting of all-or-none events. Thus, obtaining quantitative results is much less dependent on assumptions about assay efficiencies or the particulars of threshold setting. Also, there is no need to compare the results of each sample to a calibrator sample. The simplicity of analysis translates into fewer opportunities to introduce error or noise, which improves the likelihood of obtaining an accurate answer.

References

[1] B. Zimmermann, W. Holzgreve, F. Wenzel, S. Hahn, Clin. Chem. 48 (2002) 362–363.
[2] B. Bubner, K. Gase, I.T. Baldwin, BMC Biotechnol. 4 (2004) 14–23.
[3] L. Shi et al., Nat. Biotechnol. 24 (2006) 1151–1161.
[4] R.D. Canales et al., Nat. Biotechnol. 24 (2006) 1115–1122.
[5] D. Pinkel, R. Segraves, D. Sudar, S. Clark, J. Poole, D. Kowbel, C. Collins, W.-L. Kuo, C. Chen, Y. Zhai, S.H. Dairkee, B. Ljung, J.W. Gray, D.G. Albertson, Nat. Genet. 20 (1998) 207–211.
[6] N. Hosono, M. Kubo, Y. Tsuchiya, H. Sato, T. Kitamoto, S. Saito, Y. Ohnishi, Y. Nakamura, Human Mutat. 29 (2008) 182–189.
[7] T. Morrison, J. Hurley, J. Garcia, K. Yoder, A. Katz, D. Roberts, J. Cho, T. Kanigan, S.E. Ilyin, D. Horowitz, J.M. Dixon, C.J.H. Brenan, Nucleic Acids Res. 34 (2006) e123–e131.
[8] A. Dahl, M. Sultan, A. Jung, R. Schwartz, M. Lange, M. Steinwand, K.J. Livak, H. Lehrach, L. Nyarsik, Biomed. Microdevices 9 (2007) 307–314.
[9] S.L. Spurgeon, R.C. Jones, R. Ramakrishnan, PLoS ONE 3 (2008) e1662–e1668.
[10] R. Higuchi, C. Fockler, G. Dollinger, R. Watson, Nat. Biotechnol. 11 (1993) 1026–1030.
[11] P. Simmonds, P. Balfe, J.F. Peuherer, C.A. Ludlam, J.O. Bishop, A.J. Leigh Brown, J. Virol. 64 (1990) 865–872.
[12] P.J. Sykes, S.H. Neoh, M.J. Brisco, E. Hughes, J. Condon, A.A. Morley, BioTechniques 13 (1992) 444–449.
[13] B. Vogelstein, K.W. Kinzler, Proc. Natl. Acad. Sci. USA 96 (1999) 9236–9241.
[14] J. Jarvius, J. Melin, J. Göransson, J. Stenberg, S. Fredriksson, C. Gonzalez-Rey, S. Bertilsson, M. Nilsson, Nat. Methods 3 (2006) 725–727.
[15] Y.M.D. Lo, F.M.F. Lun, K.C.A. Chan, N.B.Y. Tsui, K.C. Chong, T.K. Lau, T.Y. Leung, B.C.Y. Zee, C.R. Cantor, R.W.K. Chiu, Proc. Natl. Acad. Sci. USA 104 (2007) 13116–13121.
[16] J. Qin, R.C. Jones, R. Ramakrishnan, Nucleic Acids Res. 36 (2008) e116–e123.
[17] S. Dube, J. Qin, R. Ramakrishnan, PLoS ONE 3 (2008) e2876–e2883.
[18] K.J. Livak, T.D. Schmittgen, Methods 25 (2001) 402–408.
[19] E. Schäffeler, M. Schwab, M. Eichelbaum, U.M. Zanger, Human Mutat. 22 (2003) 476–485.
[20] Y.L. Wu, S.L. Savelli, Y. Yang, B. Zhou, B.H. Rovin, D.J. Birmingham, H.N. Nagaraja, L.A. Hebert, C.Y. Yu, J. Immunol. 179 (2007) 3012–3025.
[21] J.H. Lee, J.T. Jeon, Cytogenet. Genome Res. 123 (2008) 333–342.
[22] J.R. Taylor, An Introduction to Error Analysis, University Science Books, New York, 1982.
[23] M. Kubista, J. Eliasson, M. Lennerås, S. Andersson, R. Sjöback, Eppendorf BioNews 29 (2008) 7–8.
[24] R. Sindelka, J. Jonák, R. Hands, S.A. Bustin, M. Kubista, Nucleic Acids Res. 36 (2008) 387–392.
[25] A. Tichopad, R. Kitchen, I. Riedmaier, C. Becker, A. Ståhlberg, M. Kubista, Clin. Chem. 55 (2009) 1816–1823.