A Novel, Validated Method for Absolute qPCR Quantification

Author: John SantaLucia, Jr. Ph.D.
Co-founder, CEO DNA Software, Inc.

Abstract

DNA Software's *CopyCount* software incorporates counting PCR (cPCR), a new methodology, which provides absolute values similar to digital PCR (dPCR) when using quantitative PCR (qPCR).

CopyCount determines the fluorescence from a single copy of DNA then uses this value to calculate the number of copies of DNA in an unknown sample from its fluorescence at cycle zero. Depending on the application and the amount of acceptable error, one of two different approaches can be used to measure the fluorescence from a single molecule.

Since *CopyCount* calculates the number of molecules of DNA, the need to perform a standard curve is eliminated. Rigorously validated on more than 100,000 PCR reactions, *CopyCount* provides results that are instrument and fluorophore independent, allowing comparison of results between different laboratories and meta-analysis studies of archived data sets.

CONTENTS

INTRODUCTION

The introduction of the polymerase chain reaction (PCR) in the 1980s revolutionized the field of molecular biology; PCR is now an indispensable technique for biological and clinical applications.

Shortly after the discovery of PCR, the method of quantitative PCR (qPCR) was developed[1]. Since the mechanism of PCR was not yet well understood, researchers resorted to empirical methods to analyze qPCR. The most common methods use *non-mechanism-based* curve fitting to determine a quantification cycle, Cq, which is also called the cycle threshold, C_T.

However, such methods require the laborious preparation of reference standards and acquiring a standard curve[2] to get an estimate of the absolute copy number. Alternatively, relative quantification can be determined by comparing the C_T of the sample to the C_T of a reference gene (after complex correction for "efficiency") using the $\Delta\Delta C_T$ method[3]. Best practices for utilizing these traditional methods are described in the MIQE (Minimum Information for Publication of Quantitative Real-Time PCR Experiments) guidelines[4].

In this white paper, we will show that these traditional methods introduce several misconceptions about PCR (e.g. reaction efficiency, limit of detection, and C_T for zero molecules) that compromise the quality of qPCR results. In contrast, *mechanism-based* fitting offers the advantage of more accurate quantification by qPCR using the same reagents and instrumentation but without the need to make reference standards or perform a standard curve.

EARLY ATTEMPTS AT MECHANISM-BASED FITTING

The original inventors of PCR recognized the exponential mechanism of PCR amplification[5] and Equation 1 is widely known among PCR practitioners.

$$D_N = D_0 \times 2^N \qquad \text{Equation 1}$$

N = the number of PCR cycles
D_N = the amount of DNA at cycle N
D_0 = the amount of DNA at time zero, before the PCR is performed

If Equation 1 was really true, then a researcher could just measure D_N for a variety of cycles (N) and make a graph of D_N vs. 2^N and the slope would indicate the desired D_0. However, there are several problems with applying such an approach.

The first is that we cannot directly measure the amount of DNA at each cycle. Instead, we can only measure the amount of fluorescence at each cycle, F_N, which is linearly related to the amount of DNA at each cycle by the equation:

$$F_N = m \times D_N + F_b(N) \qquad \text{Equation 2}$$

N = the number of PCR cycles
F_N = the fluorescence at cycle N
m = the slope
D_N = the amount of DNA at cycle N

$F_b(N)$ = the fluorescence background, which is a function of cycle N, and typically assumed to be linear

The slope, m, is a proportionality constant that depends upon many variables such as: quantum yield, beam-path length, sample volume, instrument geometry, quenching efficiency, beam scattering, and many more hard-to-determine effects. The values of many of these quantities vary from sample to sample, plate to plate, and instrument to instrument making it quite difficult to determine the needed D_N.

Another problem with applying Equation 1 is that the amplification process does NOT continue to be exponential for all cycles.

Equation 1 applies to a good approximation for only the early cycles of PCR (i.e. a few cycles below the C_T). Unfortunately, the fluorescence for these early cycles is dominated by fluorescence background and noise; the fluorescence signal from DNA is very small. At cycles later than the C_T, the fluorescence is above background, but Equation 1 starts to break down due to the buildup of competing amplicon re-annealing and the depletion of primers and NTPs (nucleoside triphosphates). Each cycle of PCR is becoming less efficient. In fact, it is not proper to refer to a single overall PCR efficiency.

As a result of these issues, the mechanistic fitting of qPCR to determine the absolute copy number of DNA at cycle zero has remained an unsolved problem in the field for more than 30 years.

Many researchers apply Equation 1B, with the incorrect belief that the amplification efficiency of the PCR is constant throughout the entire reaction. Equation 3 gives a much better approximation for the cycle-dependent efficiency, E_N, which accounts for amplicon re-annealing and applies up to approximately the inflection point of the PCR reaction (Figure 1).

$$D_N = D_0 \times (1 + E_N)^N \qquad \text{Equation 1B}$$

N = the number of PCR cycles
D_N = the amount of DNA at cycle N
D_0 = the amount of DNA at time zero, before the PCR is performed
E_N = the efficiency at cycle N

$$E_N = \frac{kLn\left(1+\frac{D_{N-1}}{k}\right)}{D_{N-1}} \qquad \text{Equation 3}$$

N = the number of PCR cycles
E_N = the efficiency at cycle N
D_N = the amount of DNA at cycle N

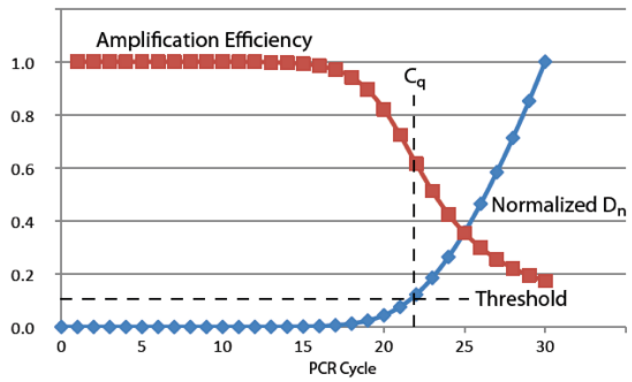


Figure 1. The amplification efficiency (red points) is cycle dependent. The DNA amplification is shown in blue. Only the first 30 cycles (i.e. before saturation) are shown.

DIGITAL PCR VERSUS QUANTITATIVE PCR

The concept of digital PCR (dPCR) was discovered in 1999 by Vogelstein and Kinzler[1]. dPCR technology allowed for the highly-accurate determination of DNA copy number without the use of reference standards.

qPCR carries out one reaction per single sample. dPCR also carries out a single reaction within a sample; however, the sample is separated into a large number of partitions (i.e. the sample is diluted such that each reaction contains 1 or 0 molecules of target DNA) and the reaction is carried out in each partition individually. This partitioning allows a more reliable collection and sensitive measurement of nucleic-acid amounts albeit at a higher cost per sample and lower throughput than qPCR.

dPCR changed our conception of Poisson sampling[6] and limit of detection, thereby opening up applications to rare-allele detection for the diagnosis of cancer and fetal aneuploidy [7,8]. In the mid-2000s, dPCR instrumentation was commercially introduced by a number of companies, including Bio-Rad, Fluidigm, Life Technologies and RainDance Technologies.

The advantages of dPCR, along with the high cost and low throughput disadvantages of the technique, inspired DNA Software. It was obvious that there was significant value if every qPCR reaction could provide an absolute quantification without requiring the user to purchase a new dPCR instrument or to change their assays.

Absolute quantification would make qPCR instrument agnostic, allowing users to compare results across different platforms. In this white paper, we describe a breakthrough in our understanding of the fundamental basis of the kinetics and thermodynamics of PCR, which enables the analysis of the shape of the qPCR curve to determine the amount of DNA in the original sample without the use of standards.

To provide absolute quantification of qPCR, all the processes going on inside a PCR reaction needed to be understood. The fundamental discovery by Boggy and Wolfe [9] that amplicon reannealing was the dominant cause of saturation in PCR got the thought process rolling. Next, the mechanisms of primer annealing, primer and other reagent depletion, were added and an initial differential equation modeled.

Finally, a way to quantify delayed-onset PCR, which is unique to analyzing the single-molecule phenomenon in qPCR, was devised and added to the model, and led to the development of counting PCR (cPCR) for absolute DNA quantification and *CopyCount* software.

ENTER COPYCOUNT

Mathematical gymnastics or mastery of DNA kinetics is not necessary to use *CopyCount* for qPCR quantification. cPCR determines the fluorescence from a single copy of DNA and then uses this value to calculate the number of copies of DNA in an unknown sample from its fluorescence at cycle zero. cPCR, and thus *CopyCount*, provides absolute values similar to dPCR, which counts copy number.

Typically, quantitation methods try to empirically match curves to quantitative values. *CopyCount* is fundamentally different and performs a first-principle analysis of the amplification curve in qPCR to generate the quantitative information, and to provide the number of molecules of DNA. Figure 2 demonstrates that the linearity continues to well below 1 molecule per well.

The use of C_T based methods has led to another incorrect concept, the "Limit of Detection" (LOD), which is commonly believed to be about 10 molecules of target DNA. The origin of this misconception is that the standard curve becomes non-linear (or undetectable for a single replicate) below about 10 molecules because there is no such thing as a C_T value for a reaction with zero molecules. In fact, PCR can detect a single molecule in a given PCR well, and if a sufficient number of replicates are acquired, PCR can detect much less than 1 molecule per well

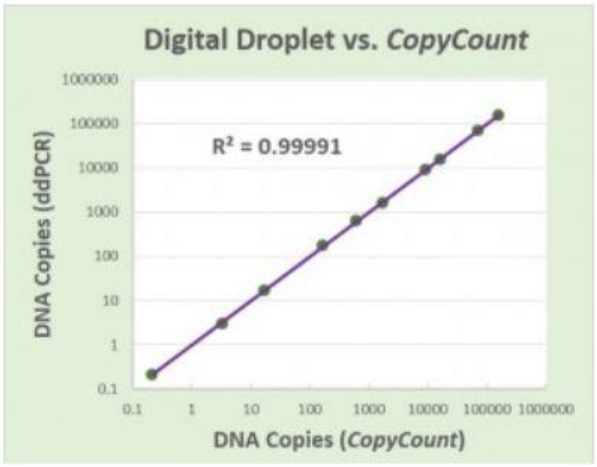


Figure 2. Comparison of the quantification from *CopyCount* with digital droplet PCR using the Bio-Rad QX-100 instrument shows excellent correlation.

COMPARABLE TO MIQE STANDARD BUT LESS LABOR

While MIQE takes the traditional, complicated, but well-proctored approach, to making qPCR quantitative, *CopyCount* takes a fundamental approach that is liberating and egalitarian.

According to Jeff Rosner, Ph.D., former Chief Technology Officer, Applied Biosystems, *CopyCount* is fundamentally different and does work. In an evaluation, the *CopyCount* calibration process was tested and assays ran with known and unknown samples, high and low replicates, as well as small and medium volume reactions using a variety of thermocyclers. In all cases, results were at least as good as or better than the standard-curve method.

Protocols generally get locked into a particular chemistry and product design, the instrument, thermo protocol, reaction volume of the reaction and so on, constraining research projects. Not only does *CopyCount* provide quantitative information the software also eliminates constraints, allowing result comparisons across platforms.

THE *COPYCOUNT* APPROACH

Consider a simple analogy of counting the number of apples in a basket (Figure 3). One could weigh all the apples in the basket, subtract the weight of the basket, and then divide by the weight of one apple. Similarly, one can count the number of copies of DNA in a sample by measuring the total fluorescence, which includes contributions from DNA and background, subtract the background, and then divide by the fluorescence of a single copy of DNA.

Figure 3. Counting the number of copies of DNA in a sample is similar to counting apples in a basket.

But there are fundamental difficulties to applying the method shown in Figure 3. First, at the beginning of PCR the total fluorescence is very similar to the background fluorescence, and so it is difficult to determine the fluorescence from DNA. At later PCR cycles (i.e. after the C_T value), the total fluorescence is much larger than then background and thus it is relatively easy to determine the fluorescence contribution from DNA.

qPCR can be used to amplify the fluorescence signal in a predictable fashion, and then the modeling incorporated into ***CopyCount*** used to analyze the high signal to noise "bend" in the qPCR curve to deduce the number of copies of DNA that are present in an unknown at cycle zero (Figure 4).

At each cycle of PCR, total fluorescence, DNA plus background, is measured. Once the background is subtracted, the fluorescence from DNA remains, and is divided by the fluorescence from a single molecule of DNA to calculate the number of DNA molecules. At early PCR cycles, total and background fluorescence are very similar in value, therefore it is the later cycles that provide the clean number of molecules used to deduce the starting concentration.

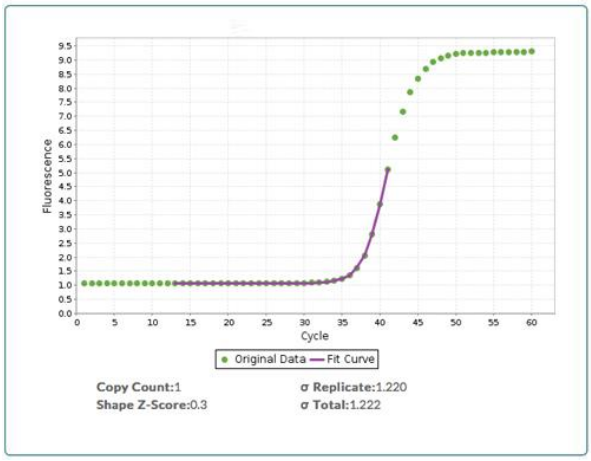


Figure 4. Output from ***CopyCount***. Raw qPCR data (green boxes) with the fit (purple line) from ***CopyCount*** for a target that has a DNA copy number = 1 at cycle zero (D_0). ***CopyCount*** fits the data in the bend of the curve.

MEASURING THE FLUORESCENCE FROM A SINGLE MOLECULE

Another difficulty is determining the quantity in the denominator in Figure 3, namely the fluorescence from a single molecule of DNA, and two different approaches can be used. The choice depends on the application.

For optimal results, the actual fluorescence from a single molecule is measured by a one-time calibration. This is accomplished by diluting the sample to ~ 1 molecule per well, which generates a Poisson distribution with approximately 36.8%, 36.8%, 18.4%, 6.1%, 1.5% of wells having 0, 1, 2, 3, 4 molecules, respectively. The measured Poisson distribution is then used to compute the fluorescence from a single molecule.

This one-time experimental calibration, which applies to any instrument, accounts for several factors that cannot be predicted, such as the amount of delayed onset, and is accurate to within 5%. If there are no changes in primer design, primer concentration, or the master mix, then this assay calibration is sufficient for all future samples.

Alternatively, DNA Software has developed fluorescence-prediction procedures. The predicted calibration is based on assumptions, such as a zero-delayed onset that does introduce some error. Some applications, such as gene-expression analysis, require that the relative amount of two genes is compared, but do not require high absolute accuracy. For these applications, *CopyCount* can predict the single-copy fluorescence without any calibration, resulting in an absolute quantification accuracy of approximately 20% and a relative quantification accuracy of 1-5% for a single well and a lower error if replicates are performed.

THE PROCESS

Once the PCR reactions are performed, PCR data are uploaded to the cloud-based ***CopyCount*****, which conducts** the analysis, including:

- Determining the range of data points to be analyzed.
- Determining which qPCR wells have a true sample versus noise.
- Carrying out replicate averaging, outlier detection, and statistical error analysis.

Absolute DNA copy number results are provided to the user in a .csv file that can be opened in Excel or a text editor. A separate analysis file is also provided that details the replicate averaging, outlier detection and statistical analyses conducted.

RIGOROUSLY VALIDATED

CopyCount has been rigorously validated on over 100,000 PCR reactions. Results are instrument and fluorophore independent, allowing comparison of results between different laboratories and meta-analysis studies of archived data sets.

NO NEED FOR A STANDARD CURVE

Since *CopyCount* calculates the number of molecules of DNA the need to perform a standard curve is eliminated. Standard curves, the traditional way to get relative quantitation in qPCR, are a series of controls and steps. They are laborious, expensive and take up valuable real estate on a multi-well plate. Plus it is difficult to evaluate if they are accurate.

According to Jane Theaker, Associate Director, Head of the IVD Technology Office Manchester, QIAGEN, standard curves can present issues. Batch-to-batch variability can occur unless standards are well calibrated and characterized, standards can become degraded unless comprehensive stability studies are performed, and technician intervention can introduce the risk of contamination. *CopyCount* has the potential to reduce the error associated with these standard curve problems. The software also makes assays more user friendly by reducing hands-on time, and leaves more wells in the plate for sample analysis thereby decreasing the overall cost.

COMPARISON OF *COPYCOUNT* TO OTHER QUANTIFICATION TECHNIQUES

	CopyCount	dPCR	Standard Curve	ΔC_T Method
Relative Quantification	✓	✓	✓	✓
Absolute Quantification	✓	✓	✓	
No Standards	✓	✓		
Not Corrupted by Low Efficiency	✓	✓		
No Gene Normalization	✓	✓		
High Accuracy	✓	✓		
Error Analysis Included	✓	✓		
No specialized Equipment	✓		✓	✓

High Throughput	✓			✓
Low Replicates	✓		✓	✓
No Dilution Series	✓			

KEY APPLICATIONS

Because cPCR uses standard qPCR instrumentation and requires fewer replicates, cPCR has higher throughput than dPCR. The technique also works with native samples without the need for sample dilution, which simplifies sample preparation.

Unlike the standard-curve method and the ΔC_T method, cPCR does not require standards, dilution series, elaborate gene normalizations, or corrections for different PCR efficiencies. Finally, cPCR includes thorough error analysis, which is useful for data interpretation in situations where data are unreliable due to poor data acquisition or poor primer design.

cPCR, and thus *CopyCount*, is appropriate for applications that require high-throughput such as viral load, fragment library quantification for next-generation sequencing, and non-invasive detection methods. It also allows meta-analysis of archived data sets.

Application	cPCR	dPCR
High-throughput Viral Titer	✓	
NGS: Fragment Library Quantification	✓	
mRNA quantification and Profiling	✓	✓
Non-invasive Detection Methods	✓	✓
Single Cell Technologies		✓
Seed Zygocity Testing	✓	
Copy Number Variation	✓	✓
Meta-Analysis of Archived Data Sets	✓	

INSTRUMENTS SUPPORTED BY *COPYCOUNT*

Life Technologies: ABI7500, ABI7900, QuantStudio 12K Flex, ViiA 7, StepOne
Roche: LightCycler, LC96, LC480
Bio-Rad: iCycler, CFX96, CFX384
QIAGEN: Rotor-Gene Q
Stratagene: Mx3000P
Abbott: m2000rt
General: Column-Layout CSV, Row-Layout CSV

COPYCOUNT USE IN METHODS DEVELOPMENT

The Borer laboratory at Syracuse University is developing high-throughput screens to simplify aptamer discovery. Aptamers are DNA/RNA molecules that have affinities for their targets similar to antibodies.

As opposed to SELEX (Systematic Evolution of Ligands by EXponential enrichment) in which the lengths of the randomized DNA or RNA region are at least 40-60 nucleotides, the Borer Acyclic Identification of Aptamers method uses shorter libraries to allow a selection process against the protein in one step, a more universal aptamer-discovery approach.

While a Ph.D. student at Syracuse University, Caitlin Miller, Ph.D. built on the acyclic identification technique evaluating the incorporation of various library structures in different ways, such as ligating adapters, and going directly to sequencing with the single-stranded library product without amplification.

Sequencing by synthesis was used. After libraries were recovered, adapters were ligated onto the 5 prime and 3 prime ends to allow the sequence to anneal to the flow cell. The exact amount of nucleic acid must be known to avoid under or overcrowding. Samples were also multiplexed; internal bar codes were used so several samples can be put on the same flow cell, ideally at the same ratio.

Dr. Miller used qPCR to determine how much single-stranded DNA was applied to the flow cell. Originally, she used the KAPA Library Quantification Kit, which uses a standard curve with six DNA standards. Reproducibility was a problem; results showed she was not getting a true representation of the samples.

The decision was made to continue to quantify with the KAPA kit and also to use *CopyCount* to compare results, and then to use the *CopyCount* determination in sequencing to see how the distribution was affected. A predicted, not experimental, calibration was used.

The ratios of the multiplexed samples were much closer using the *CopyCount* determination. In addition, the amount of high-quality data in sequencing improved, which could be attributed to applying the correct quantity of DNA to the flow cell. *CopyCount* was also able to more accurately quantify the unknown in the KAPA Library Quantification Kit standard curve then the kit itself.

In methods development sample number is low, and to run an entire standard curve for a few samples is a large expense. *CopyCount* helped save money on reagent costs, and offered more flexibility in terms of repeating experiments since a standard curve was not required.

COPYCOUNT USE IN VIRAL LOAD QUANTIFICATION

A University of Michigan study measuring cytomegalovirus (CMV) titers from plasma specimens compared quantitative results from traditional qPCR using C_T analysis and *CopyCount*[11]. The goal was to evaluate the software's feasibility for daily clinical use in a CMV plasma quantitation assay.

Fluorescence data from archived CMV assay runs were analyzed using *CopyCount* and results compared to original data using a calibration curve and C_T values. Additionally, nucleic-acid quantitative standards, calibrators traceable to international units, and patient samples were compared for linearity, precision, and accuracy.

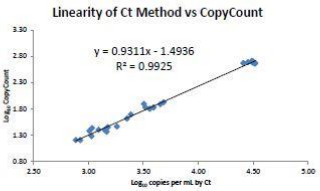
qPCR was performed on an initial target DNA sample and fluorescence data uploaded to *CopyCount* to generate an estimated copy count. Based on that estimate, a *CopyCount* 96-well experimental calibration plate was prepared, qPCR was performed and data were analyzed by *CopyCount* to evaluate acceptability.

The *CopyCount* method showed excellent linearity and comparable precision to the conventional C_T method for quantitation of both CMV from patient samples and quantitative standards (Figure 5).

Extraction efficiency must be accounted for when using either quantitation method. A correction factor for extraction efficiency was determined, and incorporated. The relationship between *CopyCount* and patient values remained linear demonstrating commutability of results between the two methods.

Precision of Extracted controls is similar for Ct vs CopyCount (n=8)

	Mean +/- SD $\log_{10}$ copies per ml		
	Low	High	Mid
Ct	3.17 ± 0.126	4.54 ± 0.0364	3.59 ± 0.144
CopyCount	3.13 ± 0.101	4.53 ± 0.0182	3.57 ± 0.16



Patient Results are commutable between Ct and CopyCount Methods

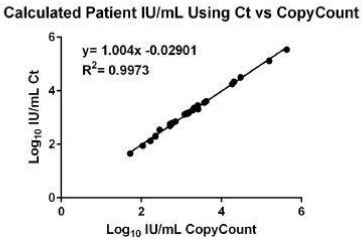


Figure 5. The *CopyCount* method showed excellent linearity and comparable precision to the conventional C_T method for quantitation of both CMV from patient samples and quantitative standards[11].

POTENTIAL CLINICAL BENEFITS

CopyCount has potential benefits for manufacturers of IVD tests. Since *CopyCount* does not require a standard curve, manufacturing costs are reduced, and plate real estate increased allowing more samples per run, decreasing the overall cost.

For example, in the conventional approach, if a test is evaluating two different targets, two standard curves are required. Including more markers for more targets in the same test kit may not be possible without converting to a smaller and less user-friendly well plate. Since there is no need to run a standard curve, *CopyCount* frees up space and potentially allows the use of more markers per test.

SUMMARY

Counting PCR (cPCR) is a quantum leap forward in the understanding of qPCR. The principle of cPCR has been incorporated into ***CopyCount***, which provides highly-reliable absolute qPCR quantification. The absolute quantification from ***CopyCount*** allows for easy comparison of results from different instruments, different samples, different targets, and different laboratories. This breakthrough in understanding the mechanism of qPCR will have significant implications for DNA-based applications.

REFERENCES

1. Vogelstein, B. and Kinzler, K.W. (1999) *Proc. Natl. Acad. Sci. USA*, **96**, 9236-9241.
2. D'haene, B., Mestdagh, P., Hellemans, J., and Vandesompele, J. (2012) *Methods Mol. Biol.*, **822**, 261-272.
3. Livak, K.J. and Schmittgen, T.D. (2001) *Methods*, **25(4)**, 402-408.
4. Bustin, S.A., Benes, V., Garson, J.A., Hellemans, J., Hugget, J., Kubista, M., Mueller, R., Nolan, T., Pfaff, M.W., Shipley, G.L., Vandesompele, J. and Wittmer, C.T. (2009) *Clinical Chemistry*, **55(4)**, 611-622.
5. Saiki, R., Gelfand, D., Stoffel, S., Scharf, S., Higuchi, R., Horn, G., Mullis, K. and Erlich, H. (1988) *Science*, **239 (4835)**, 487-491.
6. Dube, S., Qin, J. and Ramakrishnan, R. (2008) *PLoS ONE*, **3(8)**, e2876.
7. Pohl, G. and Shih, I.-M. (2004) *Expert. Rev. Mol. Diagn*. **4(1)**, 41-47.
8. Fan, H.C., Blumenfeld, Y.J, El-Sayed, Y.Y., Chueh, J. and Quake, S.R. (2009) *Am. J. Obstet. Gynecol.*, **200**, 543.e1-543.e7.
9. Boggy, G.J. and Woolf, P.J. (2010) *PLoS ONE*, **5(8)**, e12355.
10. Ruijter, J.M., Pfaffl, M.W., Zhao, S., Spiess, A.N., Boggy, G., Blom, J., Rutledge, R.G., Sisti, D., Lievens, A., De Preter, K., Derveaux, S., Hellemans, J. and Vandesompele, J. (2013) *Methods*, **59(1)**, 32-46.
11. McClellan, S., LeBar, W. and Bachman, M. (2015) "Comparison of traditional Quantitative PCR with a computer-based quantitation algorithm from plasma specimens." Poster, 31st Clinical Viral Symposium.

DNA Software
334 E. Washington St.
Ann Arbor, MI 48104
734-222-9080
info@dnasoftware.com
http://www.dnasoftware.com

Note: *CopyCount is* for research use only.