


## Increased sensitivity using real-time dPCR for detection of SARS-CoV-2

Kyra Duong<sup>1</sup>, Jiajia Ou<sup>2</sup>, Zhaoliang Li<sup>2</sup>, Zhaoqing Lv<sup>2</sup>, Hao Dong<sup>2</sup>, Tao Hu<sup>2</sup>, Yunyun Zhang<sup>2</sup>, Ava Hanna<sup>1</sup>, Skyler Gordon<sup>3</sup>, Gogce Crynen<sup>4</sup> , Steven R Head<sup>3</sup>, Phillip Ordoukhanian<sup>3</sup> & Yan Wang<sup>\*,1</sup>

<sup>1</sup>Gnomegen, 6440 Lusk Blvd, D207, San Diego, CA 92121, USA; <sup>2</sup>QuestGenomics, 12 E. Mozhou Rd, U-Park, Rm P308, Jiangning, Nanjing, Jiangsu, PR China; <sup>3</sup>Genomics Core Facility, Scripps Research, 10550 N Torrey Pines Road, La Jolla, CA, USA; <sup>4</sup>Bioinformatics Core Facility, Scripps Research, 120 Scripps Way, Jupiter, FL 33458, USA;

\*Author for correspondence: yan.wang@questgenomics.com

BioTechniques 70: 7–20 (January 2021) 10.2144/btn-2020-0133

First draft submitted: 28 August 2020; Accepted for publication: 15 October 2020; Published online: 23 November 2020

### ABSTRACT

A real-time dPCR system was developed to improve the sensitivity, specificity and quantification accuracy of end point dPCR. We compared three technologies – real-time qPCR, end point dPCR and real-time dPCR – in the context of SARS-CoV-2. Some improvement in limit of detection was obtained with end point dPCR compared with real-time qPCR, and the limit of detection was further improved with the newly developed real-time dPCR technology through removal of false-positive signals. Real-time dPCR showed increased linear dynamic range compared with end point dPCR based on quantitation from amplification curves. Real-time dPCR can improve the performance of TaqMan assays beyond real-time qPCR and end point dPCR with better sensitivity and specificity, absolute quantification and a wider linear range of detection.

### METHOD SUMMARY

A real-time dPCR instrument and assay was developed to improve detection limits and linear dynamic range of the assay, creating a more versatile and robust system for quantitating and detecting target nucleic acid sequences for infectious disease or other applications. The real-time data produced were analyzed by a simple set of heuristic functions to eliminate aberrant amplification profiles, resulting in a lower limit of detection. In addition, the fluorescent data from the entire chip can be averaged for each cycle during the amplification and used to generate a real-time amplification profile that provides quantitative information about the samples.

### KEYWORDS:

dPCR • gene expression • infectious disease diagnostics • qPCR • real-time digital PCR • real-time PCR • SARS-CoV-2

PCR-based diagnostics are used in a number of clinical applications, including detection of infectious disease pathogens [1]. The development of real-time qPCR and end point dPCR represents a significant advancement in PCR technology. Combining the amplification and detection of the target into a reaction space eliminates the need for post-amplification processing, thus simplifying workflows and reducing the risks of PCR contamination. In real-time qPCR, target quantitation is obtained using sequence-specific TaqMan<sup>®</sup> probes to measure the appearance of the target amplicon during the thermocycling process (typically in parallel with a standard curve or normalization controls); a higher initial copy number of the target results in earlier appearance (lower cycle threshold values or  $C_t$  count) of the detected amplicon [2]. In end point dPCR, absolute quantitation is achieved by performing end point PCR in up to 20,000 partitions built into a microfluidic chip or dropletized in an oil emulsion. Unlike real-time qPCR, end point dPCR does not depend on standard curves or relative threshold ( $C_t$ ) values for quantitation. In end point dPCR, the target nucleic acid is diluted to the point where at least some of the partitions contain zero copies and some partitions contain one or more copies. After end point thermocycling (often 40 cycles) the fluorescence signals in each compartment are read. Estimating the exact number of copies in the original sample is accomplished by back calculating the dilutions used in loading the sample into the partitioning system and using Poisson distribution analysis to estimate how many copies of the target molecule are contained in the ‘positive’ wells identified from the end point TaqMan signal. While real-time qPCR typically has a wider dynamic range and is more amenable to high-throughput applications [3], end point dPCR has greater accuracy and a lower detection limit and can be more precise and reproducible due to a higher tolerance to inhibitors than real-time qPCR [4,5]. End point dPCR is frequently used in applications requiring precise measurements of small differences in copy number, such as aneuploidy detection in noninvasive prenatal testing (NIPT) or low target DNA copy applications such as rare allele detection or gene fusion detection in cell-free DNA [6]. The higher sensitivity and accuracy demonstrated by end point dPCR methods in these applications, as well as its robustness to PCR inhibitors, supports its role in infectious disease detection where it is critical to detect low bacterial or viral loads [7–9]. Technology such as Fluidigm’s Digital Array exploit microfluidics to enable individually addressable partitions and can monitor real-time amplification in a dPCR experiment [10]. The Fluidigm system, while a powerful and sophisticated research tool, is neither

economical (in instrument or consumable costs) nor recommended for routine clinical applications. The microfluidic and multisample handling aspects of the system make it much more suited to research applications.

With end point dPCR, a user-determined threshold setting is applied for each end point dPCR chip based on the fluorescence intensity of each individual partition. Signals above the threshold are deemed to be positive; signals below the threshold are assigned negative status. One of the limitations of using end point dPCR with samples is the presence of a small number of false-positive signals in the end point data due to, first, instrument and fluorescent detection and imaging system-related noise, such as the location of the partitions for absorption and emission of fluorescent signal, interfering substances and distortions from image processing; and second, molecular biology noise, such as nonspecific amplifications and polymerase errors that are sequence context-driven [11,12]. Even though the number of wells giving a false-positive signal is relatively small, they can have a significant impact on determining the limit of detection (LoD) for the end point dPCR assay, especially when the number of true positive signals is also quite small [13]. As with real-time qPCR, the determination as to where to set the threshold for dPCR is somewhat subjective but can be guided by data analysis software with a manual override.

In order to apply a PCR assay developed for clinical use, a limit of blank (LoB) first needs to be determined. For example, the Thermo Fisher end point dPCR chip has ~20,000 wells, each of which yields a final end point PCR signal. If the threshold is set such that at most two positive wells are detected on 95% of known negative samples on the dPCR chips, the LoB would be two positive partitions. Any sample with equal or less than two positive partitions would be called negative. Samples with equal or greater than three positive partitions can then be called a positive result. The LoD can then be established for the lowest number of copies per sample with at least 19 out of 20 repeats resulting in a minimum of three positive wells, the level to call a sample positive in this example.

Because traditional end point dPCR only collects end point data signals, only the final end point fluorescence signal in relation to the threshold value is utilized for making the positive or negative call for each well on the chip. To address this limitation, we developed a novel real-time dPCR system that collects fluorescence data across the dPCR chip during thermocycling at user-defined cycle numbers. In this case, we determined that collecting 'real-time' data at cycles 5, 10, 15, 20, 25, 30, 35 and 39 allows us to easily distinguish most false-positive end point signals from true positive end point signals by analyzing the amplification curves revealed by the real-time data. Because we can confidently identify and remove many of the false-positive signals near the threshold value, we are able to significantly lower LoB/LoD for the real-time dPCR assay relative to strict end point dPCR data analysis.

In this study we used a SARS-CoV-2 model to evaluate and compare the sensitivity of traditional real-time qPCR and traditional end point dPCR, as well as our real-time dPCR assay for detection of SARS-CoV-2 sequence targets, using synthetic DNA and RNA spike-in, inactivated SARS-CoV-2 virus spike-in and clinical patient samples with confirmed or suspected COVID-19 (Figure 1). Our findings showed increasing sensitivity as we went from real-time qPCR to end point dPCR to real-time dPCR, with corresponding improved lower limits of detection on both analytical and clinical samples. These simple steps of using real-time data to lower detection limits provide a pathway to improving PCR-based infectious disease diagnostic assays, particularly when target copy numbers are very low. In addition, we propose a strategy to extend the linear dynamic range compared with traditional end point dPCR instruments using the real-time dPCR system's ability to also function as a real-time qPCR instrument. These two features of the real-time dPCR instrument confer greater versatility, precision and accuracy than either an end point dPCR or real-time qPCR system alone.

## Materials & methods

### Assay for evaluation & primer/probe design

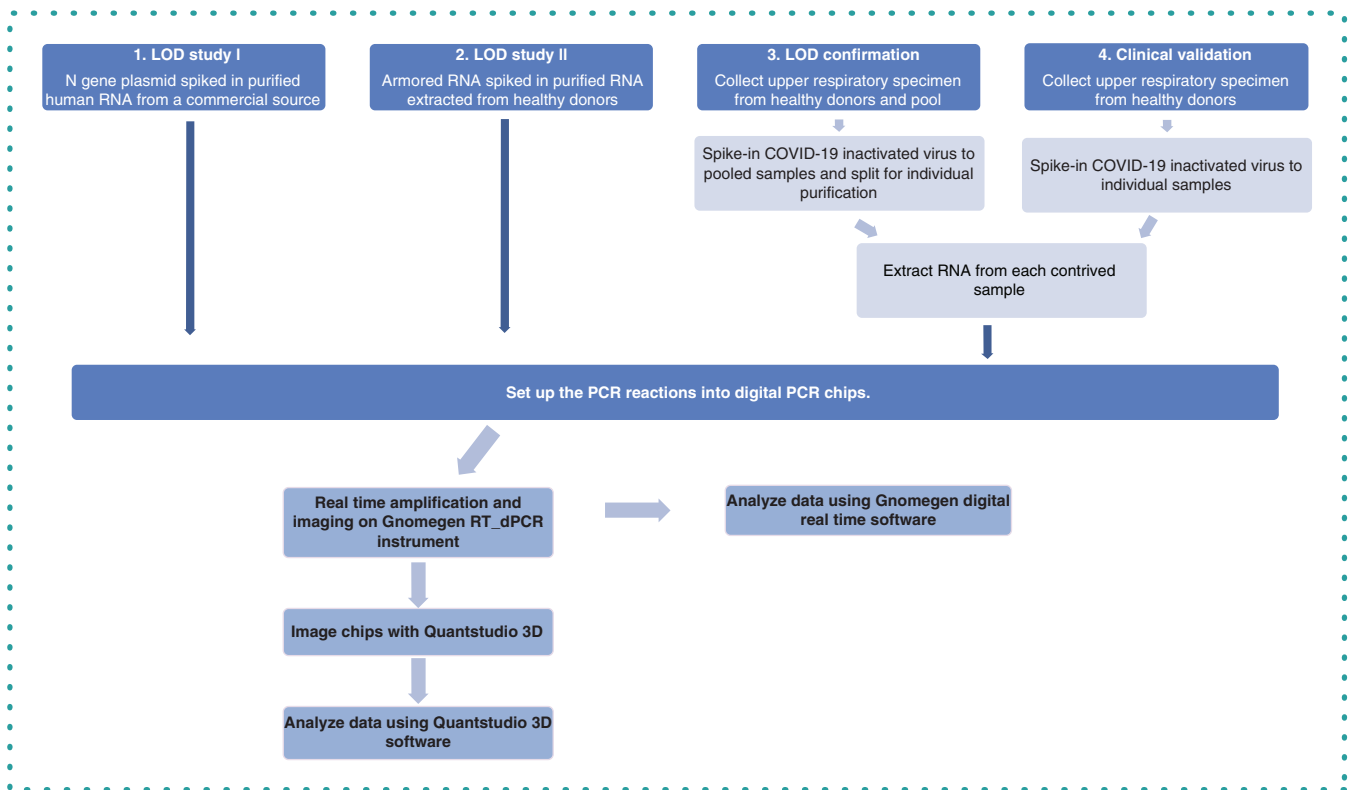
The PCR primers and TaqMan probes used in this study for real-time qPCR, end point dPCR and real-time dPCR platforms were all identical and consisted of primers and probes designed by the CDC for detection of the *N* gene of SARS-CoV-2, as described in their US FDA (EUA) [14] assay. Probes were ordered from IDT and consisted of two FAM-labeled TaqMan probes targeting independent regions of the *N* gene and one VIC-labeled probe targeting the host (human) *RNaseP* gene. Although the protocol was designed by the CDC for use in a real-time qPCR system, we also tested its performance in end point dPCR and real-time dPCR systems to evaluate the differences between the respective platforms. The real-time qPCR platform used was the QuantStudio™ 5 (Thermo Fisher, MA, USA), the end point dPCR platform used was the QuantStudio 3D (Thermo Fisher) and the real-time dPCR platform was the Gnomegen Real-Time Digital PCR Instrument (Gnomegen, CA, USA).

### Sample collection/preparation: analytical samples

Initial LoD studies were performed on characterized plasmid containing the viral *N* gene (Integrated DNA Technologies, #10006625) spiked into human genomic DNA (Promega Corporation, #G1521). Samples were created at 0, 2.5, 5, 10 and 25 copies of viral plasmid with 10 ng of human genomic DNA per 14.5- $\mu$ l reaction volume used for each Thermo Fisher end point dPCR chip. The Thermo Fisher end point dPCR chip was run on both the QuantStudio 3D Digital PCR System and the Gnomegen Real-Time Digital PCR Instrument. Additional LoD studies used inactivated SARS-CoV-2 provided by the Anhui CDC for the purposes of validating testing protocols.

### Sample collection/preparation: clinical samples

De-identified RNA samples derived from 60 upper respiratory tract specimens collected from healthy normal donors (determined to be negative for SARS-CoV-2 by real-time qPCR) were provided to us by the Anhui CDC. A contrived (control) sample was created by pooling



**Figure 1. Workflow showing sample sources and processing steps for limit of detection studies and clinical validation.** Samples for initial limit of detection estimations (LoD Study 1) were prepared using a plasmid carrying the SARS-CoV-2 *N* gene sequence. This was followed by confirmation of the limit of detection using armored RNA containing the *N* gene as well as inactivated SARS-CoV-2 virus. Contrived clinical samples were prepared by spiking inactivated SARS-CoV-2 virus into upper respiratory samples from healthy donors. For all dPCR and RT\_dPCR experiments, the Gnomegen RT\_dPCR instrument was used to collect amplification cycle data through cycle 39, at which point the chip was imaged on the QuantStudio 3D. RT\_dPCR data were analyzed using the Gnomegen digital real-time software and dPCR results were analyzed on the QuantStudio 3D software.

RNA from 12 of these samples to create a negative sample for use in establishing preliminary LoD values. Aliquots of the negative sample pool were also spiked with armored RNA containing the *N* gene (Zeesan Biotech, 2019-nCoV nucleocapsid protein *N* gene [China CDC assay]) at various concentrations to create contrived samples in triplicates at 0, 1.5, 3, 6, 12 and 24 copies viral RNA per 14.5  $\mu$ l dPCR reaction. We prepared 20 additional contrived samples similarly at the tentative LoD and tested them by end point dPCR. The tentative LoD was then confirmed if at least 19 out of 20 repeats were determined to be positive. We ran real-time qPCR assays in parallel on the same purified RNA samples to determine the LoD by real-time qPCR.

Once the tentative LoD was determined for end point dPCR, the 60 individual samples were prepared to include 30 negatives (no spiked-in pseudovirus), 20 samples with pseudovirus spiked in at  $1.5 \times$  LoD and 10 samples with pseudovirus spiked in at  $4 \times$  LoD. All 60 samples were tested following the manufacturer's instructions for the Gnomegen COVID-19 RT-Digital PCR detection kit (Gnomegen, CA, USA, #CV0202).

### Real-time qPCR

The commercial real-time qPCR kit (Bioperfectus Technologies Co, Ltd., #JC10223-1N) was used for the detection. The 20- $\mu$ l reaction contained 7.5  $\mu$ l of nucleic acid amplification reaction buffer, 5  $\mu$ l of enzyme mix, 4  $\mu$ l of COVID-19 reaction buffer and 5  $\mu$ l of RNA. Thermal cycling was performed at 50°C for 10 min for reverse transcription, followed by 97°C for 1 min (predenaturation) and then 45 cycles of 97°C for 5 s (denaturation) and 58°C for 30 s in the Roche Light Cycler® 480 real-time PCR system (Roche, Basel, Switzerland).

### End point dPCR

TaqMan assays were run on end point dPCR chips (Thermo Fisher, #A26317) prepared following the manufacturer's recommended protocols for the QuantStudio 3D Digital PCR System using the Gnomegen COVID-19 RT-PCR Detection Kit. Briefly, 14.5  $\mu$ l of reaction mix is required for each reaction, comprised of 7.25  $\mu$ l of dPCR master mix (Thermo Fisher, #A26358), 0.20  $\mu$ l Superscript II (Thermo Fisher, #18064014), 0.20  $\mu$ l RNaseOUT (Thermo Fisher, #10777019), 0.725  $\mu$ l COVID-19 assay, 2.125  $\mu$ l molecular grade nuclease-free water and 4  $\mu$ l RNA sample. Following the protocols for the end point dPCR System, the 14.5  $\mu$ l of reaction mix was loaded on the chips using the QuantStudio 3D Digital PCR Chip Loader (Thermo Fisher, #4482592), then thermocycled in ProFlex™ 2  $\times$  Flat Block thermal cycler

Table 1. Algorithm used for real-time data to flag atypical amplification curves for removal in real-time dPCR experiments.

	False positive to be removed	Standard (which will be removed)
1	Fluorescent signal increases but then decreases before completion of thermocycling	FAM20/FAM15 <0.95 or FAM25/FAM20 <0.95 or FAM30/FAM25 <0.95 or FAM35/FAM30 <0.95
2	Fluorescent signal comes up too late	(FAM39-FAM35)/(FAM39-FAM15) >0.3
3	Fluorescent signal starts high in early cycles or remains relatively flat during thermocycling	FAM20 > MEAN(FAM20) + 6*SD(FAM20) or (FAM35-FAM25)/25 <0.1
4	Fluorescent signal shows minimal increase between cycle 25 and cycle 35	(FAM30-FAM25)/(FAM39-FAM15) <0.33
5	Fluorescent signal shows too much increase between cycle 10 and cycle 15	0.92 < FAM15/FAM10 <1.08

SD: Standard deviation.

(Thermo Fisher, #4484078) under the following cycling protocol: step 1, 42°C for 20 min (reverse transcription); step 2, 96°C for 10 min (DNA polymerase activation, denaturation) and 60°C for 2 min (annealing); step 3, 39 cycles of 98°C for 30 s (denaturation) and 60°C for 2 min (annealing); step 4, 20°C (cooling) for infinite hold. Finally, the chips were imaged by QuantStudio 3D Digital PCR Instrument (Thermo Fisher, #4489084) and analyzed using the QuantStudio 3D Analysis Suite Cloud Software (version 3.1.4-PRC-build1) to measure the concentration of *N* gene (FAM probes) and *RNase P* gene (VIC probe), respectively.

## Real-time dPCR

The reaction mixes used in end point dPCR were also used in real-time dPCR and thermocycled on the Gnomegen Real-Time Digital PCR Instrument (Gnomegen, #INS1). The use of the instrument and SARS-CoV-2 assay has been granted EUA approval by the FDA [15]. The instrument is configured for chip-based dPCR with a filter system that has two decoupled excitation filters and six decoupled emission filters. The instrument collects real-time raw fluorescence data following the extension step of the PCR cycle. Thermal cycling was performed at 42°C for 10 min for reverse transcription, followed by 39 cycles of 95°C for 30 s (denaturation), 60°C for 2 min (annealing). During thermocycling, the chips were imaged every five cycles after the extension step and at the end point (cycle 39). After thermocycling, the chips were removed and end point-imaged in the Thermo Fisher QuantStudio 3D Chip reader (Thermo Fisher, #A29154) following the manufacturer's recommended protocol.

## Positive-negative discrimination of real-time dPCR

Real-time curves generated for critical positive/negative discrimination were analyzed by observation of atypical amplification profiles that conformed to an algorithm as described in Table 1. From this algorithm, false-positive data points were removed from further analysis.

## Linear dynamic range analysis of real-time dPCR

Fluorescent signals were detected and monitored at cycles 5, 10, 15, 20, 25, 30, 35 and 39 from each dPCR chip on the Gnomegen Real-Time Digital PCR instrument, and amplification profiles were generated from data in the VIC channel. Approximately 18,000 wells were averaged into a single 'averaged value' and plotted as a function of cycle for each of the total RNA inputs (0.1, 1, 10 and 100 ng). This averaged VIC fluorescent signal was plotted along the Y-axis and cycles along the X-axis. Data points for each amplification profile were fitted to a symmetrical sigmoidal curve (4PL) with the formula: (generated from curve fit software available from mycurvefit.com):

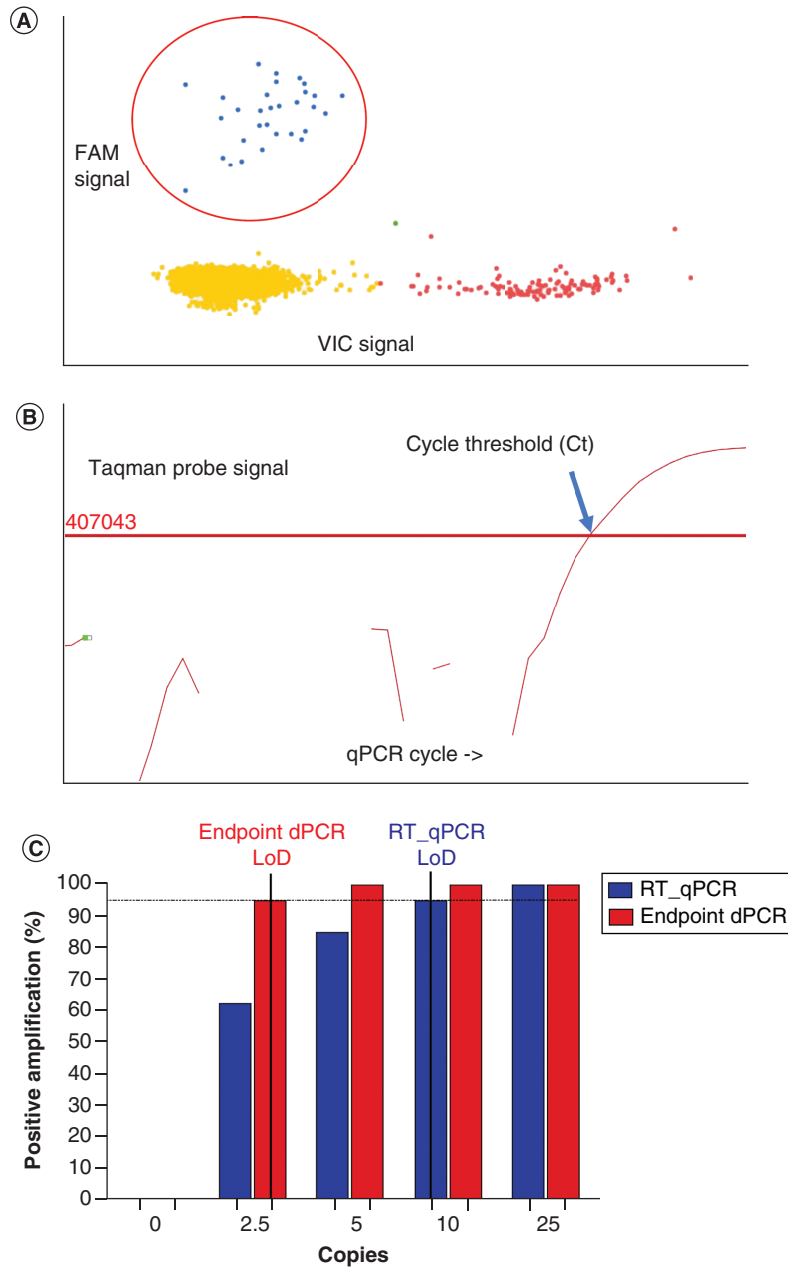
$$y = d + \frac{a - d}{1 + \left(\frac{x}{c}\right)^b}$$

where *y* is the fluorescence and parameters *a*, *b*, *c*, *d* and *x* correspond to minimum asymptote, slope, inflection point, maximum asymptote and cycle number respectively. Each data series produced curve-fitting algorithms, and with a threshold set at 1.3 fluorescent signal, they were used to generate *C<sub>t</sub>* values that were then plotted against the log (input)(pg).

## Results & discussion

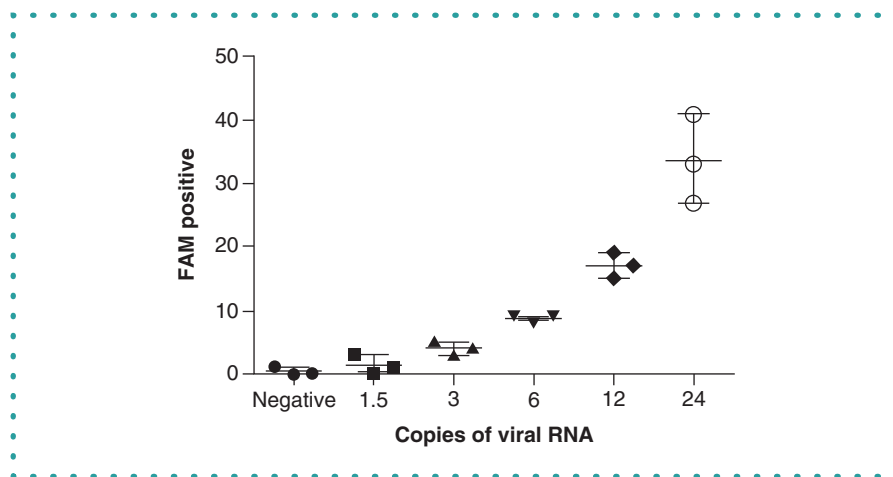
### End point dPCR versus real-time qPCR: limit of detection

Our initial goal was to estimate the LoD for real-time qPCR and end point dPCR using viral *N* gene sequence encoded in a plasmid vector mixed with human gDNA. The strategy for LoD determination is to test various concentrations of the target sequence in a uniform background such that the only variable is the target concentration. Figure 2A & B show the data formats generated from real-time qPCR and end point dPCR, respectively. Positive signals in real-time qPCR present as the appearance of TaqMan probe signal and the cycle at which the amplification curve reaches the 'cycle threshold' (*C<sub>t</sub>* or *C<sub>q</sub>* value), which is inversely proportional to the target



**Figure 2.** Graphs of preliminary limit of detection for real-time qPCR and end point dPCR. The positive results of (A) end point dPCR and (B) real-time qPCR for a reference standard created using synthesized plasmid containing the SARS-CoV-2 *N* gene mixed with human genomic DNA at ten copies of plasmid with 10 ng of human gDNA per reaction. The blue plots in the red circle (A) represent positive signals of SARS-CoV-2 *N* gene for the end point dPCR. The sample is positive when the  $C_t$  value of the real-time qPCR amplification curve (B) crosses with the threshold (the blue arrow) at  $\leq 37$  cycles. The limits of detection for real-time qPCR and end point dPCR were tested with 20 replicates per concentration at 0, 2.5, 5, 10 and 20 copies of target plasmid containing the SARS-CoV-2 *N* gene sequence mixed with 10 ng human gDNA in each reaction. (C) The preliminary limit of detection for end point dPCR is 2.5 copies, while that for real-time qPCR is 10 copies using a 95% positive threshold rate (dotted line is 95%).

copy concentration in the sample. In end point dPCR, positive wells are those wells on the chip that are generating sufficient TaqMan probe signal (in this case a FAM-labeled probe). We tested 0, 2.5, 5, 10 and 20 copies of target plasmid containing the SARS-CoV-2 *N* gene sequence mixed with 10 ng human gDNA in each reaction; both real-time qPCR and end point dPCR assays were performed on these same reference standards. For real-time qPCR, we required a positive result to have a  $C_t$  value  $\leq 37$  cycles, in line with the CDC's recommended cutoffs for positive  $C_t$  values for SARS-CoV-2 testing. For end point dPCR, a positive/negative threshold was set such



**Figure 3. End point dPCR results of SARS-CoV-2 tests from contrived clinical samples to evaluate the limit of detection.** 12 healthy donor respiratory samples were pooled and spiked in 0, 1.5, 3, 6, 12 and 24 copies (per 14.5- $\mu$ l reaction) of the pseudovirus carrying the *N* gene sequence of SARS-CoV-2 with three replicates for each concentration. Mean values and coefficients of variation for each concentration were 0.33 (141.42%), 1.33 (93.54%), 4.00 (20.41%), 8.67 (5.44%), 17 (9.61%) and 33.67 (17.03%), respectively. Copies of viral RNA values were plotted on the X-axis versus measured values of FAM positive signals on the Y-axis using Graph Pad Prism (GraphPad, CA, USA).

that all negative samples contained  $\leq 2$  positive signals while positive samples contained  $> 3$  positive signals on the end point dPCR chip.

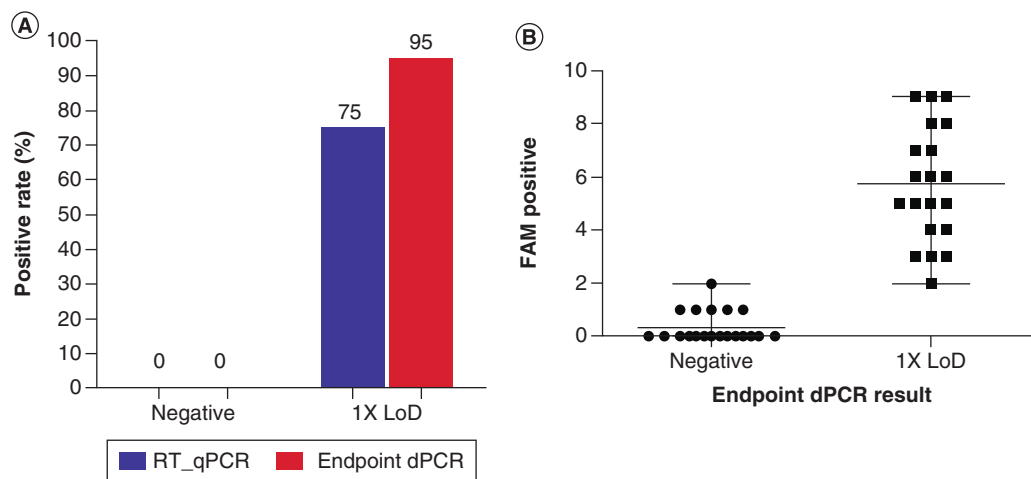
With these threshold settings, each concentration was tested with 20 replicates. The preliminary LoD results are shown in Figure 2C. For both real-time qPCR and end point dPCR, all negative samples were called negative (i.e., there were no false positives). At 2.5 copies per reaction, 19/20 samples were called positive by end point dPCR, whereas only 62.5% were called positive by real-time qPCR. At 5 copies per reaction, end point dPCR called 100% of samples positive, while real-time qPCR calls reached 95% positive only at 10 copies per reaction and 100% positive at 20 copies per reaction. Based on these results, we determined a preliminary LoD of 2.5 copies per reaction for end point dPCR and 10 copies per reaction for real-time qPCR.

Next we evaluated the LoD for real-time qPCR and end point dPCR in the context of actual human respiratory samples collected from healthy donors. For this experiment we pooled respiratory samples from 12 healthy donors, spiked with 0, 1.5, 3, 6, 12 and 24 copies (per 14.5- $\mu$ l reaction) of the pseudovirus carrying the *N* gene sequence of SARS-CoV-2. By pooling the donor specimens, we eliminated sample-related variability to focus on target detection (LoD) in the context of a biological specimen. Three replicates were prepared for each concentration, and identical reference sample sets were run on real-time qPCR and end point dPCR systems. Figure 3 shows the results for end point dPCR. Using the same criteria for calling samples positive, we set the threshold such that the presumptive LoD in this sample set was three target copies per end point dPCR chip. With the LoD set at three copies, we tested 20 replicates by end point dPCR and real-time qPCR (Figure 4A). The results show that 19/20 samples tested positive by end point dPCR, while only 15/20 were positive by real-time qPCR. There were no false positives in either system. After plotting the number of positive signals for each sample/dPCR chip in the negative and  $1 \times$  LoD groups (Figure 4B), we can see the distribution of positive signal values for each sample. All negative samples contained  $\leq 2$  positive signals, while 19/20 positive samples contained  $\geq 3$  positive signals on the dPCR chip.

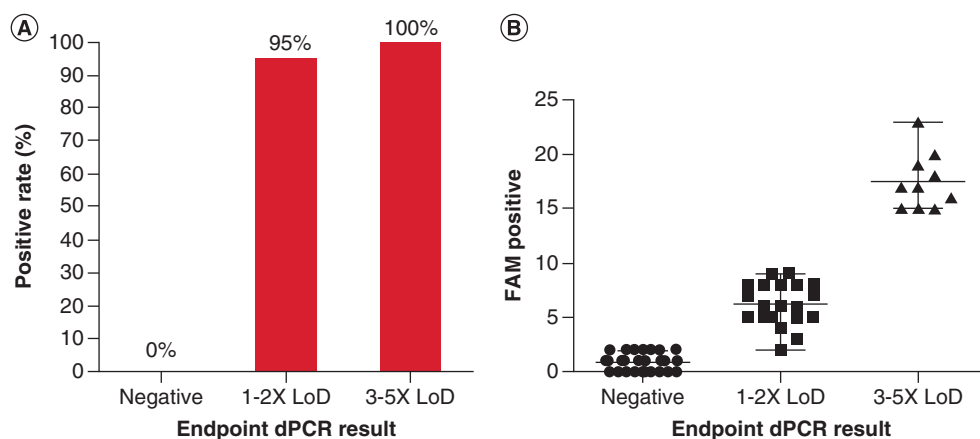
This established the functional LoD value for the pseudovirus containing *N* gene target at three copies per reaction – a lower LoD than we were able to obtain by real-time qPCR. Next we tested a set of contrived clinical samples consisting of independent donor samples (no longer using pooled samples) to more closely reflect a real-world testing situation. We obtained 60 confirmed negative clinical respiratory samples from healthy donors. In 20 of these samples, we spiked in the plasmid carrying the *N* gene sequence of SARS-CoV-2 at 4.5 copies per reaction ( $1.5 \times$  LoD) and 10 samples at 12 copies per reaction ( $4 \times$  LoD). The human *RNase P* reference target was also evaluated; all 60 samples had *RNase P* positive value signals ( $\geq 60$ ; data not shown). Figure 5A shows the end point dPCR results of samples: 95% of the  $1.5 \times$  LoD samples (4.5 copies per reaction) and 100% of the  $4 \times$  LoD samples (12 copies per reaction) gave a positive call. The distribution of positive FAM wells in the end point dPCR chips is shown in Figure 5B. Here, again, we can see 1–2 false-positive FAM signals in a substantial portion of the negative samples, presenting a potential lower cutoff for the LoD of true positive samples.

### Real-time dPCR: linear dynamic range

Using the Gnomegen Real-Time Digital PCR instrument, we performed real-time dPCR thermocycling and captured fluorescent images at cycles 5, 10, 15, 20, 25, 30, 35 and 39. The captured data were used to assess amplification of the TaqMan signal in 'real time' on the dPCR chip. This merging of real-time quantitative PCR data with dPCR data in a single platform presents a novel method with which



**Figure 4. Graphs and data of limit of detection for real-time qPCR and end point dPCR.** (A) The detection of the end point dPCR system compared with the real-time qPCR system on the pooled negative samples (20 replicates) and  $1 \times$  preliminary LoD (3 copies per reaction) contrived clinical samples (20 replicates). All negative samples for real-time qPCR and end point dPCR were called negative (no false positives detected). For the  $1 \times$  LoD samples, 95% (19/20) were positive as detected by end point dPCR, while only 75% (15/20) were called positive by real-time qPCR. (B) Expected sample types of negative and  $1 \times$  LoD samples were plotted on the X-axis versus measured values of FAM positive signals on the Y-axis, targeting the *N* gene of SARS-CoV-2 in each sample. All 20 negative samples showed  $\leq 2$  FAM positive signals, while 19/20  $1 \times$  LoD samples contained  $\geq 3$  positive signals and only a single sample had 2 positive signals. Mean and coefficient of variation values for negative and  $1 \times$  LoD samples were 0.55 (121.63%) and 3.95 (28.28%), respectively. LoD: Limit of detection.



**Figure 5. End point dPCR assay results on contrived clinical sample consisting of independent donor samples.** (A) End point dPCR system detection result of negative samples (30 replicates),  $1.5 \times$  LoD samples (4.5 copies per reaction, 20 replicates) and  $4 \times$  LoD samples (12 copies per reaction, 10 replicates). Expected sample types were plotted on the X-axis versus measured values of FAM positive signals on the Y-axis, targeting the *N* gene of SARS-CoV-2. All the negative samples were called as negative; the positive rates of  $1.5 \times$  LoD and  $4 \times$  LoD samples were 95 and 100%, respectively. (B) FAM positive signals distribution of these samples by end point dPCR. Expected sample types were plotted on the X-axis versus measured values of FAM positive signals on the Y-axis, targeting the *N* gene of SARS-CoV-2. All 30 negative samples were true negative but contain 1–2 false-positive FAM signals; the  $1.5 \times$  LoD samples (20 samples) produced 19 (95%) true positives with 1 false negative. All samples at the  $4 \times$  LoD concentration were called as positive. Mean and coefficient of variation values for negative,  $1-2 \times$  LoD and  $3-5 \times$  LoD samples were 0.90 (82.90%), 6.20 (30.86%) and 17.50 (14.06%), respectively. LoD: Limit of detection.

Table 2. Comparison of end point dPCR results with real-time dPCR results on 12 different negative clinical samples.

Sample	FAM+ (real-time dPCR)	FAM+ (end point dPCR)
S1	0	2
S2	0	2
S3	0	1
S4	0	2
S5	0	1
S6	0	2
S7	0	0
S8	0	0
S9	0	2
S10	0	2
S11	0	2
S12	0	8
Mean	0	2.00
SD	0	1.96
CV	NA	97.89%

CV: Coefficient of variation; SD: Standard deviation.

to view and observe the PCR amplification process. We decided to focus on two main criteria that could be improved upon by such a process: first, whether we could capitalize on the real-time qPCR data generated in this digital dPCR format to expand the dynamic range of sample input for this dPCR system compared with end point dPCR; and secondly, whether we could investigate the false-positive FAM signals found in negative test samples that present a lower boundary on LoD values in positive test samples.

We first addressed the question about the linear dynamic range in end point dPCR and real-time dPCR systems. In standard end point dPCR, one molecule of target nucleic acid per well is ideally desired, so the linear dynamic range becomes limited on the low end by subsampling error and on the high end by Poisson statistics, such that more than one target molecule can end up in the same well. At the high or low end, the end point dPCR can no longer provide any quantitative interpretation, and the relationship between data points becomes nonlinear. This creates an optimal window for a linear signal response, outside of which the response is much less linear. The real-time dPCR instrument collects real-time data for every single well of the dPCR chip (typically around 20,000); an example of a portion of the raw data is depicted in Figure 6A. To evaluate whether we could use the real-time data collected on the Gnomegen dPCR system to expand the linear dynamic range of quantitation for the RNA target sample compared with end point dPCR, we generated amplification profiles from the real-time fluorescent signals, averaged across all wells on the dPCR chip, from the *RNase P* TaqMan probe for the total RNA inputs of 0.1, 1, 10 and 100 ng per reaction (Figure 6B). The data were fitted to a sigmoidal curve and a threshold was set in the exponential phase for all curves to determine a  $C_t$  value. Next, to look at the linearity of the data, we plotted the estimated copy number based on the end point dPCR data only (taken from cycle 39; Figure 6C) or plotted the real-time data (corresponding  $C_t$  values for each curve) against the input amounts (Figure 6D). The real-time dPCR data demonstrated better linearity and consequently higher  $R^2$  values (Figure 6D). The end point dPCR data (Figure 6C) exhibited a nonlinear relationship due to the fact that some of the data points lay outside the optimal window for end point dPCR quantitation ( $R^2 = 0.8805$ ), although when the fluorescent signals from all wells were analyzed as real-time data, the relationship became much more linear within the dynamic range ( $R^2 = 0.9958$ ), outperforming end point dPCR by being able to measure concentrations that are saturated by end point dPCR. The use of the real-time data improves the useful linear dynamic range of the real-time dPCR instrument.

### Real-time dPCR: LoD

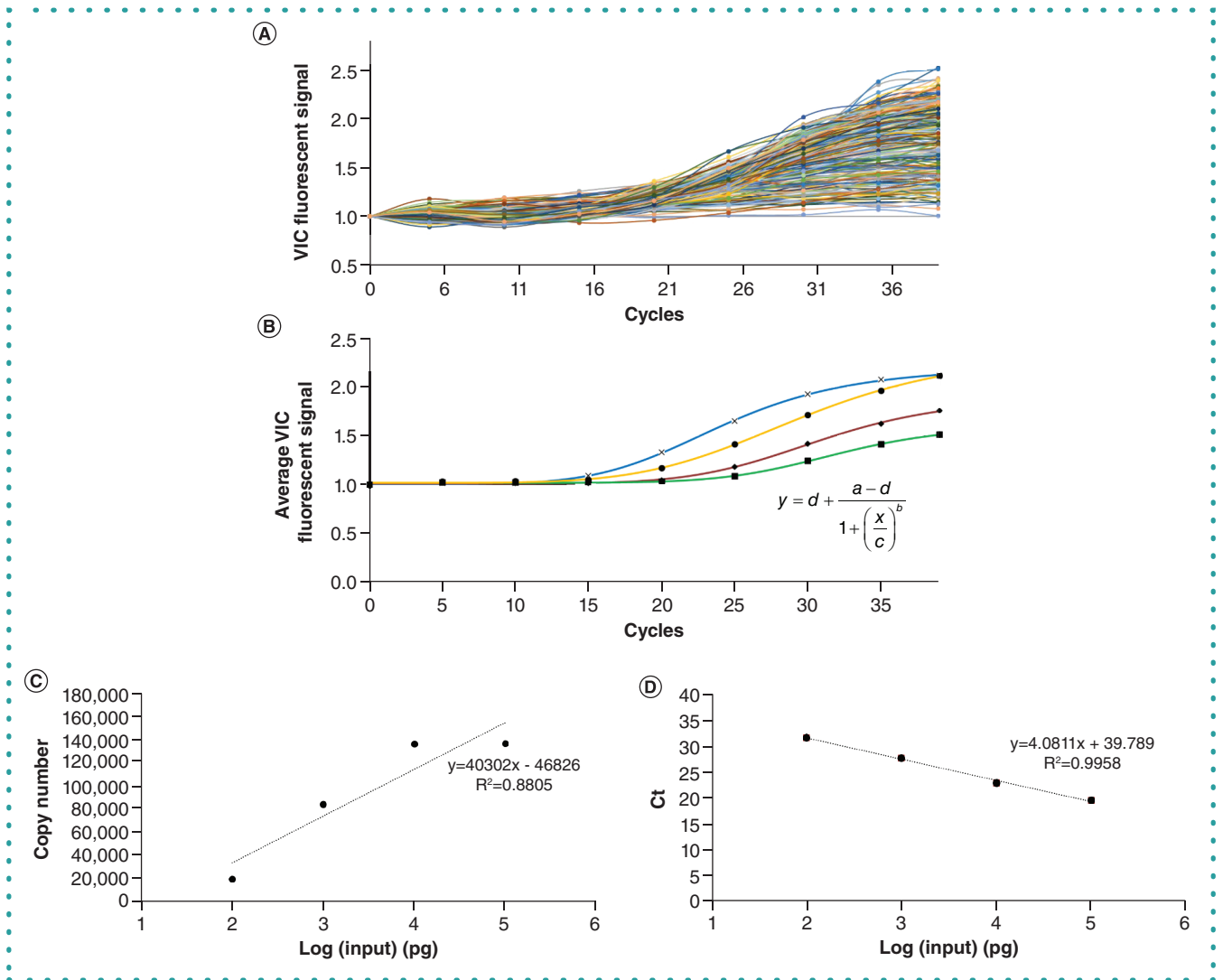
Next we decided to explore the nature of the false-positive signals found in the negative samples. If these could be objectively characterized as fluorescence signals from a source other than specific PCR amplification, they could be removed from further analysis, thereby removing this lower boundary on testing LoD. Figure 7 provides two examples of using real-time curves to remove false-positive signals: Figure 7A & B show real-time curves from a positive sample, and Figure 7C & D show real-time curves from a negative sample. Figure 7A & C show examples of real-time curves from positive signals on a dPCR chip based on end point threshold setting; Figure 7B & D show true positive real-time curves after removal of atypical amplification curves deemed to be false positives. It is our observation that every dPCR chip has a small number of aberrant data points that can be identified by the atypical amplification profiles. Identification and removal of these false-positive data points allows the threshold setting to be lowered, ultimately improving the LoD for positive samples by allowing the possible inclusion of data points that would have been eliminated with a higher threshold setting. For identification of atypical amplification curves, we used the algorithm shown in Table 1.

Table 2 displays a comparison of end point dPCR results and the real-time dPCR results on 12 different negative clinical samples.



dPCR chips were amplified and real-time imaged on the Gnomegen RT dPCR instrument but were also read in the Thermo Fisher end point dPCR instrument for comparison. As shown, using real-time dPCR to analyze the samples, none of the 12 chips showed positive signals when our algorithm was applied to remove false-positive signals from each sample. However, when the dPCR chips were scanned in the Thermo Fisher end point dPCR instrument, the processed data showed a range of positive signals (typically 1 or 2). This set the LoB at a level of two positive signals.

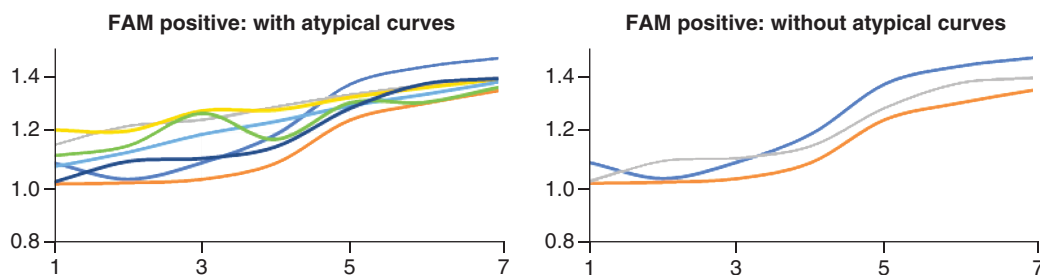
Next we analyzed a set of samples to compare traditional end point dPCR with dPCR plus real-time data correction to remove false positives. Sample types include 1–2 × LoD contrived clinical samples with inactivated SARS-CoV-2 spiked in at ten genomic copies per



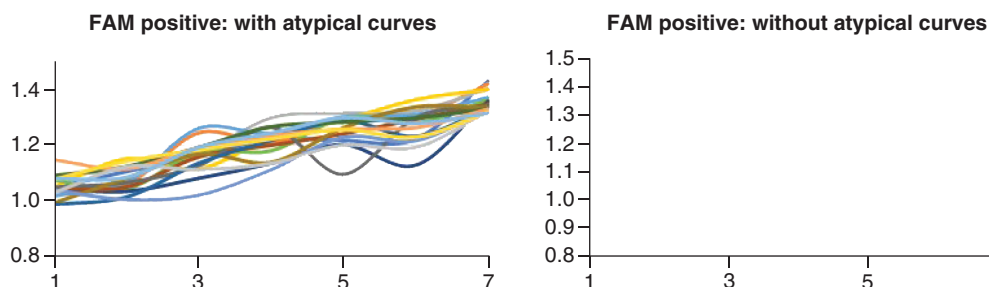
**Figure 6. Linear dynamic range comparison of Thermo Fisher end point dPCR versus Gnomegen real-time dPCR. (A)** Example of amplification profiles of the raw data from the first 254 of approximately 18,000 total wells of 100 ng RNA input sample. **(B)** Averaged amplification profiles of human *RNase P* (reference) generated in the VIC channel on the Gnomegen real-time dPCR instrument. Fluorescent signal was detected and monitored at cycles 5, 10, 15, 20, 25, 30, 35 and 39 from each dPCR chip. Fluorescent signals from approximately 18,000 wells per chip were averaged, and the average values

were plotted as a function of cycle for the different total RNA input amounts: 0.1 (—■), 1 (—◆), 10 (—●) and 100 ng (—×). Data points were averaged from two chips for each of the following inputs: 0.1, 1 and 100 ng; the second 10-ng chip had a failure due to technical issues, so only one chip was used. The averaged VIC fluorescent signal (*RNase P*) was plotted along the Y-axis, and cycles of amplification along the X-axis. Data points for each amplification profile were fitted to a sigmoidal curve with the formula  $y = d + \frac{(a - d)}{[1 + (x/c)^{1/b}]}$  (as depicted), with the algorithm for each curve fitting the data with an  $R^2$  of 0.9982, 0.9993, 0.9995 and 0.9997, respectively. A threshold of 1.3 fluorescent signal was used in the algorithm (this intersected all plots during the exponential phase) for each curve to calculate a  $C_t$  value for each of the RNA input concentrations. **(C)** The Thermo Fisher QuantStudio 3D end point dPCR system was used to generate copy number data for the four RNA input amounts (0.1, 1, 10 and 100 ng); copy number for each input amount was plotted against the log (input amount)(pg);  $R^2$ , slope and intercept are displayed on the graph. **(D)**  $C_t$  values for the four RNA input amounts (0.1, 1, 10 and 100 ng), generated from the algorithms of the graph on the Gnomegen real-time dPCR instrument (B), were plotted versus the log (input amount)(pg);  $R^2$ , slope and intercept are displayed on the graph.

Example 1: C0GZFG (May batch data\_1XLOD\_s41)



Example 2: D043BT (negative\_s2)



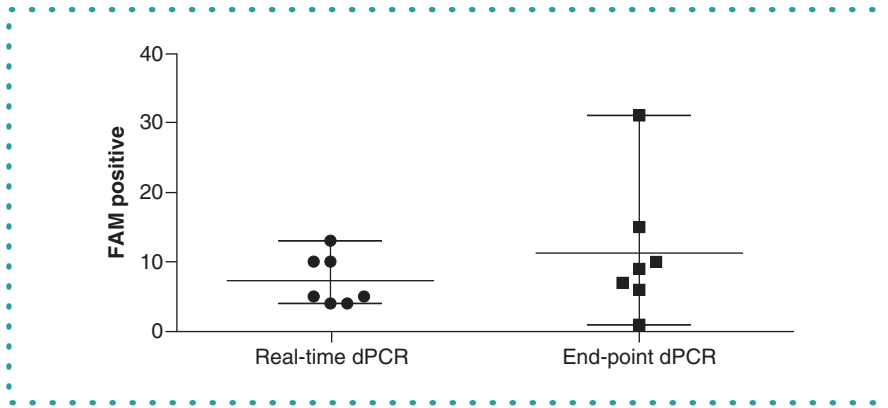
**Figure 7. Identification and removal of false-positive data points allows the threshold setting to be lowered and ultimately improves the limit of detection for positive samples. (A & B) Examples of real-time curves from a positive sample. (C & D) Examples of real-time curves from a negative sample. A & C show examples of real-time curves from positive signals on a dPCR chip based on end point threshold setting; B & D show true positive real-time curves after removal of atypical amplification curves deemed to be false positives. The numbers plotted on the X-axis represent different time points during PCR amplification. The positive signal value is shown on the Y-axis.**

**Table 3. Comparison of real-time dPCR results with end point dPCR results on 1–2 × limit of detection contrived clinical samples (March batch).**

Sample	FAM+ (real-time dPCR)	FAM+ (end point dPCR)
S13	4	9
S14	10	6
S15	5	7
S16	4	10
S17	13	31
S18	10	1
S19	5	15
Mean	7.29	11.29
SD	3.37	8.96
CV	46.23%	79.35%

CV: Coefficient of variation; SD: Standard deviation.

reaction before RNA purification. The same digital PCR chips were run in the Gnomagen real-time Digital PCR instrument with real-time imaging data collected and then read in the Thermo Fisher QuantStudio 3D Digital PCR reader for end point reading. Traditional end point dPCR analysis is shown as the number of 'FAM Positive End point' wells for each dPCR chip. Table 3 shows the number of 'real-time corrected FAM positive' wells remaining after removal of potentially positive wells with atypical amplification curves. This shows the comparison of end point dPCR results with real-time dPCR results in these 1–2 × LoD contrived clinical samples. As shown, after removing potential false-positive signals, real-time dPCR analysis resulted in a better distribution of positive FAM signals (4–13) versus end point dPCR (1–31) on the same exact dPCR chips, where the only difference was that the real-time data were used to correct the end point data. This means that real-time dPCR not only removes false-positive wells effectively, but also allows the user to relax the threshold and include more positive wells that would be removed with a more stringent threshold (Figure 8).



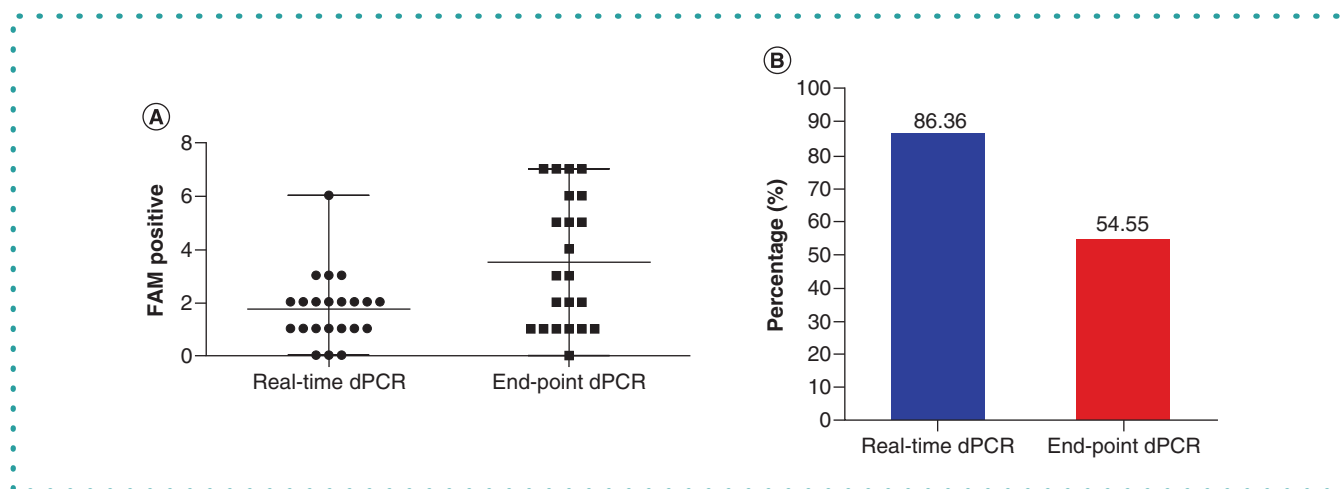
**Figure 8.** Distribution of FAM positive numbers of 1–2 × limit of detection contrived clinical samples in a single batch for real-time dPCR and end point dPCR. After potential false-positive signals were removed, real-time dPCR analysis resulted in better distribution of the positive FAM numbers (4–13) than end point dPCR (1–31). The numbers of FAM positive wells per sample are plotted on the Y-axis.

**Table 4.** Retesting of ‘aged’ 1–2 × limit of detection contrived clinical samples by real-time dPCR and end point dPCR (May batch).

Sample	Read by real-time dPCR			Read by end point dPCR		
	FAM+	Judgment	Positive rate	FAM+	Judgment	Positive rate
S20	2	+	86.36%	2	-	54.55%
S21	0	-		0	-	
S22	2	+		6	+	
S23	1	+		7	+	
S24	2	+		2	-	
S25	1	+		1	-	
S26	2	+		7	+	
S27	2	+		6	+	
S28	1	+		1	-	
S29	2	+		5	+	
S30	1	+		1	-	
S31	1	+		5	+	
S32	1	+		1	-	
S33	6	+		7	+	
S34	3	+		3	+	
S35	3	+		3	+	
S36	0	-		2	-	
S37	0	-		4	+	
S38	2	+		1	-	
S39	2	+		5	+	
S40	1	+		1	-	
S41	3	+		7	+	

The ‘aged’ samples were a set of 1–2 × limit of detection clinical samples (similar in composition to those in Table 3) that had been aged for 2 months at -80°C.

Furthermore, we tested this approach on another set of 1–2 × LoD contrived clinical samples (similar in composition to the samples in Table 3) which had been ‘aged’ for 2 months at -80°C. We assumed that for these samples the viral target copies would be at the lowest level of detection before aging, and any sample degradation would result in a reduction in the ability to call these contrived samples as positive. When we compared traditional end point dPCR with real-time dPCR and used the real-time data analysis correction to remove false positives (Table 4) we found that of the 22 1 × LoD samples, end point dPCR was able to call 12 (~55%) as positive relative to the 4 negative samples. However, when the real-time data correction was applied, the number of 1 × LoD samples called as positive increased to 19 (~86%), demonstrating the improved sensitivity of real-time dPCR. Additionally, the real-time dPCR positive signals formed a much tighter pattern than in the end point dPCR data (Figure 9), suggesting better quantification, accuracy and precision.



**Figure 9.** Distribution of FAM positive numbers and consistency rate of test results by real-time dPCR and end point dPCR on the ‘aged’ 1–2 × limit of detection contrived clinical samples. (A) After removing potential false-positive signals, real-time dPCR analysis resulted in a tighter pattern of positive FAM data points than end point dPCR. The FAM positive numbers are plotted on the Y-axis. (B) The sensitivity of end point dPCR was 12/22 (54.55%); when real-time data correction was applied to remove false positives, this increased to 19/22 (86.36%).

### Real-time qPCR, end point dPCR & real-time dPCR

Real-time qPCR and end point dPCR are both highly versatile, well-characterized methods for generating amplification data that can be used for quantitation and detection of target sequences down to extremely low levels. Each method has its own list of areas and applications where one technique might be more advantageous than the other. Real-time qPCR has a wider dynamic range, making it more versatile, and is more amenable to high-throughput operations; end point dPCR demonstrates better resistance to inhibitors, resulting in greater accuracy and reproducibility, as long as the data fall within an optimal concentration window. It was our goal to build a novel instrument that would harness both techniques, gaining the advantages of both methods in a single assay, and consequently allowing us to eliminate some of the noise associated with false-positive detections and potentially widen the dynamic range compared with end point dPCR alone.

At present, real-time qPCR is widely used in the detection of infectious diseases, including HIV and SARS-CoV-2 [13,16]. In this study we wanted to compare the diagnostic abilities and dynamic range for the detection of SARS-CoV-2 by real-time qPCR, end point dPCR and real-time dPCR. Our data show that the use of end point dPCR demonstrates a higher sensitivity with a lower LoD relative to real-time qPCR. However, our novel method of real-time dPCR was able to further improve the sensitivity of detection. In this approach, fluorescent signals are collected during the dPCR thermocycling process, enabling the identification and removal of false-positive data points by their atypical amplification curve profiles. In addition, we made use of the real-time data to widen the linear dynamic range of dPCR methodology. Typically, end point dPCR is limited to four orders of magnitude of dynamic range, whereas real-time qPCR commands around seven orders of magnitude [17]; using our real-time dPCR assay, we were able to generate real-time qPCR-like data from a dPCR instrument, expanding the versatility of this instrument.

In traditional end point dPCR, several factors can affect the accuracy of quantitation, such as the user-defined positive/negative threshold, which is also used similarly in the real-time dPCR experiments. This threshold is used to define whether a well is deemed positive or negative based on the intensity of the fluorescent signal for each specific target sequence on the dPCR chip. In end point dPCR, the threshold is set based on the end point image and used to define the final set of positive wells. However, in real-time dPCR, amplification curves are generated for every single ‘tentative’ positive signal and the amplification profiles are used to evaluate each one. Because the number of atypical amplification curves is usually small (~1–15 per chip) relative to the total number of wells (~20,000), their removal is only significant when the true positive signal counts are very low; for example, near the LoD of the assay. These spurious data points are generated by abnormalities on the dPCR chip, such as autofluorescence caused by dust or other contaminants on the chip or in an emulsion. A second issue is that compartments producing unintended ‘abnormally shaped’ amplification profiles as a result of nonspecific amplification typically generate much higher fluorescent signals. Both abnormalities can result in unintended false-positive interpretations. However, through the analysis and use of real-time data, these spurious data points were removed, improving the data twofold by first removing spurious data points from the data signals and second allowing the detection threshold to be lowered, increasing the number of ‘true’ positive signals that could be observed and counted.

To provide a means of objectively identifying false-positive signals, we defined a set of amplification profile parameters that were used to remove data points that did not conform to a typical doubling-type amplification profile, as would be expected for real-time qPCR data. The rules we defined specifically for our dataset may not be universally applicable to all real-time dPCR experiments, but a number of amplification curve characteristics identified here are likely to be useful in evaluating any real-time dPCR assay, including:

- A lack of increase in amplification levels across cycles – wells that start out high early in cycling, possibly from a dust particle or debris with high autofluorescence;
- Fluctuating amplification curves, where a signal may be affected by neighboring signals or dark areas;
- A nonspecific amplification profile that behaves in an uncharacteristic way.

We used a set of simple heuristics to define our atypical amplification profiles; it seems likely that in future work, machine learning approaches could be effectively applied to larger data sets to flag atypical profiles for removal. As different TaqMan assays have different signal-to-noise ratios and different shaped amplification curves, it is not realistic to have a one-size-fits-all algorithm. For different assays to be developed for clinical use, it is completely feasible to develop an algorithm once the following criteria are met:

- Primers and probes are tested and optimized
- A significant number of samples are run at different levels of input DNA to obtain a profile of a 'standard' amplification curve
- The algorithm to be used is tested in different batches of the primers and probes of interest on the sample types that would be encountered in a clinical setting.

## Future perspective

End point dPCR has provided a huge improvement in precision and reproducibility, but not necessarily sensitivity, over real-time qPCR. It also has achieved absolute quantification without an external reference and demonstrated greater resistance to inhibitors [4,5]. However, end point dPCR can be limited by a lower linear dynamic range, is less amenable to high-throughput applications and has been shown to generate false-positive data points that can throw into question the results for some low-detection applications [10]. It was our goal to design and develop an instrument that could combine real-time qPCR and end point dPCR in a single assay, so it would have the attributes of both systems. In this way, the real-time qPCR data could be used to remove false positives and widen the linear dynamic range for the assay. In removing false positives, we used the real-time qPCR data for each well and identified atypical amplification profiles. In doing this, we would be removing both 'instrument-related fluorescent detection' type noise and also some of the 'molecular biology' noise, like nonspecific amplification. Potential polymerase errors generated in the early cycles of amplification at the site of the probe, which can also contribute to false positives [13], would not be addressed by the new instrument, but it is not inconceivable that algorithms may be designed to detect even these to some degree in future work. A full analysis of the real-time data for every single well on the dPCR chip was outside the scope of this study; however, it is possible that in future work, a full analysis of that type might elicit greater advantages to the system that were not demonstrated here.

The Gnomegen COVID-19 detection assay was used as an example here to demonstrate the utility of this real-time dPCR technology. We were able to improve detection sensitivity of SARS-Cov-2 targets with better precision and for lower copy numbers. Real-time dPCR is a promising new technology for applications that require high sensitivity, such as circulating tumor DNA and NIPT in blood samples, or detection of low viral loads associated with infectious diseases. It can also be used for applications that require accurate quantification at extremely low levels; for example, to detect amplification of genes such as *HER2* and *cMET*, which are of significant interest in cancer research, or for NIPT analysis in the blood for cell-free circulating DNA. Real-time dPCR instruments can meet these technical needs for both research and clinical applications.

## Author contributions

The manuscript was written by K Duong, P Ordoukhanian, SR Head and Y Wang. J Ou, G Crynen, Z Lv and H Dong provided data analysis. A Hanna, Z Li, S Wong and S Gordon processed PCR samples.

## Financial & competing interests disclosure

S Head and P Ordoukhanian are paid consultants for Gnomegen, LLC. The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

No writing assistance was utilized in the production of this manuscript.

## Open access

This work is licensed under the Attribution-NonCommercial-NoDerivatives 4.0 Unported License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>

## References

1. Yang S, Rothman RE. PCR-based diagnostics for infectious diseases: uses, limitations, and future applications in acute-care settings. *Lancet Infect. Dis.* 4(6), 337–348 (2004).
2. Panzarin V, Patarnello P, Mori A *et al.* Development and validation of a real-time TaqMan PCR assay for the detection of betanodavirus in clinical specimens. *Arch. Virol.* 155(8), 1193–1203 (2010).
3. Wong ML, Medrano JF. Real-time PCR for mRNA quantitation. *BioTechniques* 39, 75–85 (2005).

4. Taylor SC, Laperriere G, Germain H. Droplet Digital PCR versus qPCR for gene expression analysis with low abundant targets: from variable nonsense to publication quality data. *Sci. Rep.* 7, 2409 (2017).
5. Hindson CM, Chevillet JR, Briggs HA *et al.* Absolute quantification by droplet digital PCR versus analog real-time PCR. *Nat. Methods* 10(10), 1003–1005 (2013).
6. Zhou R, Cai Y, Li Z *et al.* A digital PCR assay development to detect *EGFR* T790M mutation in NSCLC patients. *Front. Lab. Med.* 2(3), 89–96 (2018).
7. Hayden RT, Gu Z, Ingersoll J *et al.* Comparison of Droplet Digital PCR to real-time PCR for quantitative detection of cytomegalovirus. *J. Clin. Microbiol.* 51(2), 540–546 (2013).
8. Zhao Y, Xia Q, Yin Y, Wang Z. Comparison of Droplet Digital PCR and quantitative PCR assays for quantitative detection of *Xanthomonas citri* subsp. *citri*. *PLoS ONE* 11(7), e0159004 (2016).
9. Li H, Bai R, Zhao Z *et al.* Application of droplet digital PCR to detect the pathogens of infectious diseases. *Biosci. Rep.* 38(6), BSR20181170 (2018).
10. Azuara D, Ginesta M, Gausachs M *et al.* Nanofluidic digital PCR for *KRAS* mutation detection and quantification. *Gastrointest. Cancer Clin. Chem.* 58(9), 1332–1341 (2012).
11. Rowlands V, Rutkowski AJ, Meuser E *et al.* Optimisation of robust singleplex and multiplex droplet digital PCR assays for high confidence mutation detection in circulating tumour DNA. *Sci. Rep.* 9, 12620 (2019).
12. Milbury CA, Zhong Q, Lin J *et al.* Determining lower limits of detection of digital PCR assays for cancer-related gene mutations. *Biomol. Detect. Quantif.* 1(1), 8–22 (2014).
13. Kiselina M, Pasternak AO, De Spiegelaere W *et al.* Comparison of Droplet Digital PCR and seminested real-time PCR for quantification of cell-associated HIV-1 RNA. *PLoS ONE* 9(1), e85999 (2014).
14. CDC 2019–Novel Coronavirus (2019-nCoV) real-time RT-PCR diagnostic panel (2020). <https://www.fda.gov/media/134922/download>
15. COVID-19 RT-Digital PCR Detection Kit Instructions for Use (2020). <https://www.fda.gov/media/136738/download>
16. Li D, Wang D, Dong J *et al.* False-negative results of real-time reverse-transcriptase polymerase chain reaction for Severe Acute Respiratory Syndrome Coronavirus 2: role of deep-learning-based CT diagnosis and insights from two cases. *Korean J. Radiol.* 21(4), 505–508 (2020).
17. Gerwyn MJ, Busby E, Garson JA *et al.* Digital PCR dynamic range is approaching that of real-time quantitative PCR. *Biomol. Detect. Quantif.* 10, 31–33 (2016).