Virtual partition digital PCR for high precision chromosomal counting applications

Lucien Jacky[1]*, Dominic Yurk[1,2], John Alvarado[1], Bryan Leatham[1], Jerrod Schwartz[1], Chris MacDonald[1], Aditya Rajagopal[1,2,3]

1 ChromaCode Inc., 2330 Faraday Ave Suite 100, Carlsbad, CA 92008, USA
2 Department of Electrical Engineering, California Institute of Technology, Pasadena, CA 91125, USA
3 Department of Biomedical Engineering, University of Southern California, Los Angeles, CA 90089, USA

* ljacky@chromacode.com

Abstract

Digital PCR (dPCR) is the gold standard analytical platform for rapid high precision quantification of genomic fragments. However, current dPCR assays are generally limited to monitoring 1-2 analytes per sample, thereby limiting the platform's ability to address some clinical applications that require the simultaneous monitoring of 20 – 50 analytes per sample. Here we present Virtual Partition dPCR (VPdPCR), a novel analysis methodology enabling the detection of 10 or more target regions per color channel using conventional dPCR hardware and workflow. Furthermore, VPdPCR enables dPCR instruments to overcome upper quantitation limits caused by partitioning error. While traditional dPCR analysis establishes a single threshold to separate negative and positive partitions, VPdPCR establishes multiple thresholds to identify the number of unique targets present in each positive droplet based on fluorescent intensity. Each physical partition is then divided into a series of virtual partitions, and the resulting increase in partition count substantially decreases partitioning error. We present both a theoretical analysis of the advantages of VPdPCR and an experimental demonstration in the form of a 20-plex assay for non-invasive fetal aneuploidy testing. This demonstration assay – tested on 432 samples contrived from sheared cell-line DNA at multiple input concentrations and simulated fractions of euploid or trisomy-21 "fetal" DNA – is analyzed using both traditional dPCR thresholding and VPdPCR. VPdPCR analysis significantly lowers variance of chromosome ratio across replicates and increases the accuracy of trisomy identification when compared to traditional dPCR, yielding >98% single-well sensitivity and specificity. VPdPCR has substantial promise for increasing the utility of dPCR in applications requiring ultra-high-precision quantitation.

Introduction

In many clinical diagnostic applications, it is essential to not only detect the presence of a nucleic acid target, but to also measure its concentration. This is commonly done with the quantitative polymerase chain reaction (qPCR), which calculates concentration based on the number of PCR cycles needed for a sample to reach a certain signal threshold. This method benefits from widespread instrument deployment and a simple

workflow, but requires running standards for calibration purposes and is only effective at establishing target concentration within a factor of 2 [1–3].

An alternative approach to target quantitation is digital PCR (dPCR). dPCR divides a sample into tens of thousands of individual partitions prior to amplification. After amplification, the partitions that generate an amplification signal are counted to produce measurements of target concentration. When compared to qPCR, dPCR improves both precision and low-copy sensitivity, while providing absolute quantitation without the need for calibrating standards [1–4]. Due to its advantages, dPCR has shown increasing utility in a number of diagnostic applications, including absolute quantification of viral load, analysis of circulating DNA, gene and microRNA expression, and analysis of gene copy number variation [5–9].

The most fundamental limit to the precision of dPCR quantitation is Sampling Variance; if the true mean number of target copies across many sample replicates is N, the standard deviation of the number of copies truly present in each replicate will be $\sqrt{N}$ [10] (Figure 1a). For applications that require high precision at low input copy numbers, this level of variance can be unacceptably high. The number of effective sample copies can be increased by multiplexing, i.e. designing multiple assays specific to different regions of the target. However, as the number of effective sample copies passes the number of physical partitions the partitions become oversaturated, leading to high levels of Partitioning Variance [10]. This tradeoff, illustrated in Figure 1b, has limited the usefulness of multiplexing for decreasing Sampling Variance in single-well dPCR assays [5].

The statistical analysis underlying the red line in Figure 1b assumes traditional dPCR analysis, which uses a single amplitude threshold to classify each partition as either negative for all targets or positive for at least one target. One way around this Partitioning Variance limitation is to apply enhanced assay design to generate multiple distinct signal clusters corresponding to different combinations of target regions in the partitions. Historically this process has proven difficult, limiting such multiplexing applications to 2 or 3 targets per optical color channel [11–15]. However, this difficulty can be overcome by applying High Definition PCR (HDPCR$^{\text{TM}}$), a recent innovation in qPCR technology, which has been used to expand the multiplexing capacity of qPCR instruments [16]. In this paper we describe Virtual Partition dPCR (VPdPCR), a method which leverages HDPCR and a novel analysis technique to enable significantly higher levels of multiplexing on existing dPCR instruments using standard TaqMan chemistries. We present both a theoretical analysis of the advantages of VPdPCR and an experimental demonstration of its capabilities in the form of a single-well, 20-plex assay for detection of fetal chromosomal aneuploidies.

Virtual Partition dPCR

VPdPCR combines multiple TaqMan® assays that are designed to detect multiple distinct regions of the desired target, in this case a particular chromosome. The TaqMan probes for all target regions on a given chromosome are labeled with the same fluorophore and quencher pair, and they are all titrated to generate the same fluorescence intensity if the target region is present. These probes generate signals that add linearly in combination; if a single target region generates fluorescence intensity I when it is present in a partition, a partition with n distinct target regions will have fluorescence intensity $n * I$. The number of target regions presents in each partition can thus be inferred solely from the signal intensity measurement of the partition.

In traditional dPCR analysis, a single signal intensity threshold is drawn to separate partitions positive for 1 or more target regions from those partitions negative for all target regions. VPdPCR changes the readout of a digital partition from a 2 state system to a $T + 1$-state system by drawing T different intensity thresholds, where T is

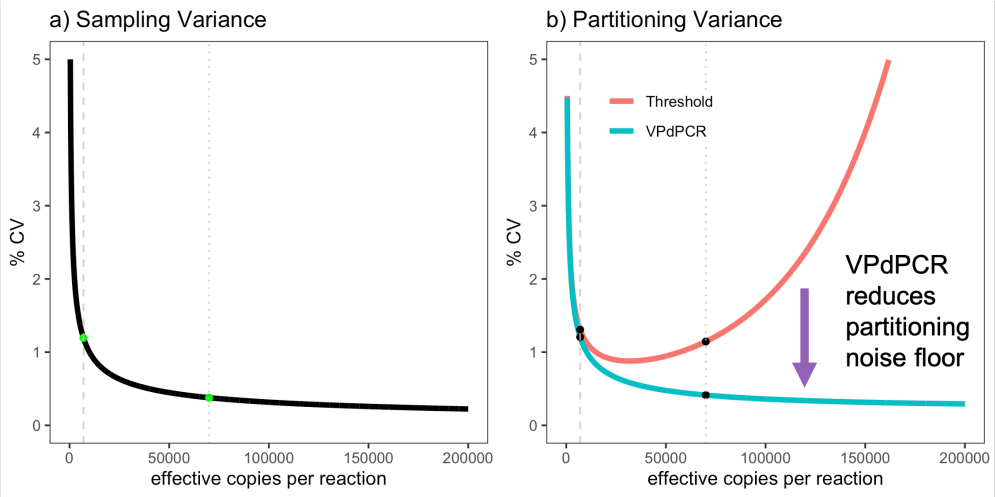

Fig 1. Digital PCR mathematical noise floor The coeffecient of variation (CV) of calculated copies per reaction across a range of expected copies in sample, assuming a digital PCR reaction with 20,000 physical partitions. The dashed grey line represents 7,000 copies per reaction with only one assay per target, while the dotted grey line represents 70,000 effective copies corresponding to assays for 10 unique target regions on each of 7,000 target copies. a) Sampling Variation: The standard deviation due to sampling error is defined as the square-root of the mean expected copies in a sample. This error is independent of the analysis method. Increasing the number of assays per target effectively increases the expected number of copies and decreases the Sampling Variance. b) Partitioning Variation: The standard deviation due to partitioning error is dependent on the number of partitions a sample is divided into. The VPdPCR assay (blue) described in this paper increases the effective number of partitions 10-fold, significantly reducing the mathematical noise floor due to partitioning error when compared with the tradition threshold method (red).

the number of target regions and 1 is added for the negative state (Figure 2). Each partition is divided into T "virtual partitions", enabling higher allowable target concentrations without oversaturation by increasing the number of effective partitions by a factor of T. Creating these virtual partitions significantly reduces the Partitioning Variance which usually occurs as the occupancy of the physical partitions approaches 100% (Figure 1b, blue curve).

This manuscript presents a demonstration assay capable of detecting 10 unique target regions per channel, with one channel devoted to targets from chromosome 18 and the other dedicated to targets from chromosome 21 (Figure 3). By increasing the effective target concentration, VPdPCR drives down Sampling Variance while avoiding high Partitioning Variance caused by oversaturation (Figure 1). This allows for consistent detection of very small differences in abundance ratio between the chromosomes. To demonstrate the utility of the technique, we apply it to the problem of differentiating simulated cell-free DNA (cfDNA) samples with and without small fractions of trisomy 21 DNA.

Non-Invasive Fetal Aneuploidy Testing

Screening for fetal aneuploidy in expectant mothers is one of the most common forms of prenatal diagnostics in the world [17], and is traditionally performed using methods such as chorionic villus sampling or amniocentesis. While these tests still represent a

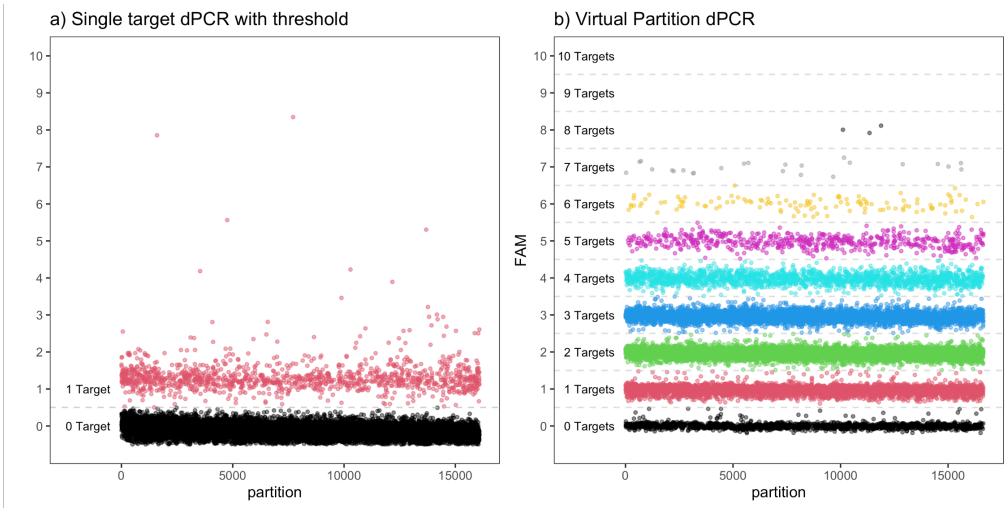

Fig 2. Evolving Multiplex Digital PCR a) Traditional singleplex dPCR is a robust method using a simple threshold but is limited by significant variation at low input and high input due to sampling and Partitioning Variance respectively. b) By combining VPdPCR with HDPCR we're able to divide each partition into many bins to determine the copies per virtual partition. This substantially reduces Partitioning Variance by expanding the number of effective partitions.

gold standard for accuracy, their invasive nature limits their application to high-risk populations [18]. The discovery of fetal cell-free DNA (cfDNA) circulating in maternal blood [19] opened the door to non-invasive prenatal testing (NIPT) for fetal aneuploidy by counting chromosomal copies; if a fetus is triploid for a given chromosome, the number of copies of that chromosome in fetal cfDNA should be 50% higher than all other copy numbers. Modern fetal aneuploidy tests are often performed via next-generation sequencing or microarray tests which require expensive equipment and consumables as well as complicated multi-day workflows, largely relegating them to centralized laboratories and driving up costs.

dPCR provides lower cost, lower complexity, and higher throughput when compared to NGS or microarray tests, making it a desirable modality for fetal aneuploidy screening. However, no single well dPCR-based assay for NIPT has come to market [20]. The primary reason is limited precision; for a euploid mother and an aneuploid fetus with a trisomy only the fetal portion of the cfDNA will show an excess in chromosomal copies, and the fraction of fetal cfDNA derived from a maternal blood or plasma sample can be as low as 4% [21]. This results in only a 2% excess of the fetal trisomy chromosome amongst the whole cfDNA sample, and standard single target dPCR assays are unable to distinguish maternal from fetal DNA. As a result, attempting to consistently measure the excess of the fetal trisomy chromosome using single-plex dPCR proves especially difficult due to Sampling Variance.

Multiple groups have attempted to use multiplexing to bypass this Sampling Variance problem [22–24]. However, in order to avoid oversaturation, these assays must increase the number of effective partitions by either splitting each sample across 8 or more wells [22, 24] or using specialized (and now-discontinued) platforms that can generate millions of partitions per sample [23]. Both of these approaches increase cost and decrease throughput. We instead apply VPdPCR to increase the number of effective partitions, thereby substantially improving accuracy of quantitation at higher target concentrations than was previously possible in a single-well assay. While the presented results are only intended as a proof of concept, they establish the power of

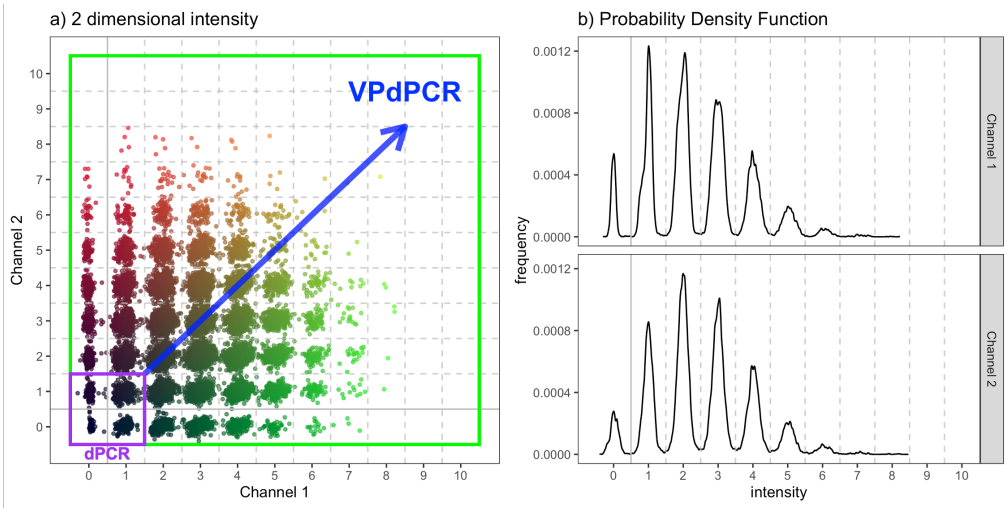

Fig 3. Virtual Partition dPCR data a) Two dimensional plot of experimental VPdPCR data with 7000 haploid genomic copies in 20,000 physical partitions. The assay was designed with 10 independent target regions per chromosome, with chr18 target regions in channel 1 and chr21 target regions in channel 2. Thanks to HDPCR, distinct point clusters can be discerned containing droplets with different numbers of target regions. The color of the individual points represents the intensity of channel 1 and channel 2 as the Green and Red component of RGB values respectively. The purple box encompasses the partitions that can be analyzed with standard dPCR thresholding, while the green box encompasses the additional virtual partitions which can be interrogated in a 10 target per channel VPdPCR assay. b) Probability density function (PDF) plots for channel 1 and channel 2 of the experimental data from the same reaction. Each channel's PDF is analyzed independently to calculate copies of each chromosome. In both plots the solid grey line represents the traditional threshold between positive and negative physical partitions, while the dashed grey lines separate the virtual partitions differentiated by signal intensity.

VPdPCR as a potential foundation for future ultra-high-precision quantitative assays.

Glossary

- Target: the whole nucleic acid molecule of which the concentration is being interrogated. Examples: a whole genome, an individual chromosome, a particular RNA transcript.
- Target Region: a sub-sequence within the complete Target sequence which is detected by a unique assay. Example: the template region of a Target for a PCR detection assay.
- Partition: one of many independent physically separate PCR reactions into which a sample is equally divided in dPCR. Examples: A individual droplet or microfluidic well.
- Virtual partition: Expanded partitions derived from the signal amplitude when multiple Target Region assays are leveled to produce the same signal intensity. Example: in an assay with 10 Target Regions to interrogate a single Target, every physical partition is divided into 10 virtual partitions using the VPdPCR method.

- Classification: Analyzing a partition and determining how many positive targets it contains based on its signal amplitude. Examples: identifying a partition as positive or negative in traditional dPCR, or placing it into a bin level 0 through T in VPdPCR.

- Positive Partition Count: the total number of positive partitions identified after classification. Examples: The sum of positive physical partitions in traditional dPCR or the sum of positive virtual partitions in VPdPCR.

- Target Region Copies: the computed total number of target regions present in a sample after obtaining a Positive Partition Count and applying Poisson statistics. Example: Chr18 Copies is the imputed total number of target regions from chromosome 18 present in the original sample.

- Sampling Variance: The variation the number of targets which actually end up in a reaction due the sub-sampling of a larger population. This has larger effect on low concentration sample accuracy due to the standard deviation being $\sqrt{m}$ where m is the expected number of targets [25].

- Partitioning Variance: The variance attributed to the distribution of targets between the partitions. At high concentrations the number of empty partitions the standard deviation $\sqrt{E(1-E)/n}$ where E is the proportion of negative partitions and n is the number of partitions [10].

Results and Discussion

We contrived 432 cell line-based DNA samples, each consisting of a mixture of "maternal" euploid DNA from a wild type cell line and simulated "fetal" DNA from either a trisomy 21 cell line or a different euploid cell line. Half of the samples contained 7000 total copies of chr18 and half of the samples contained 3500 copies – both of which are representative of DNA concentrations in a typical cell-free DNA extraction from expectant women [26]. Within each of these sets, 36 samples each had simulated aneuploid trisomy 21 fetal fractions of 0%, 5%, 10%, and 20%, and 36 samples each had simulated euploid fetal fractions of 5% and 100%. Two of the 20% simulated aneuploid fetal fractions wells at 7000 input copies had failures in the droplet reader and results were discarded. Summary results are shown below in Table 1 with all euploid samples condensed into one row, and these euploid samples are broken out by cell line composition in Table 2. An unexpected result is that even in the purely euploid samples the chr21 counts are 4.9% higher than chr18 counts on average. There are multiple possible explanations for this, including off-target amplifications, a duplicated target region on chr21, or chr18 target regions being more susceptible to shearing during sample preparation. Fortunately, this excess is consistent across all euploid sample compositions and the difference in ratios between this baseline and other experimental conditions (ΔRatio in the table) scales as expected with aneuploid fraction, indicating that this is a consistent offset across all samples. Thus, when classifying samples to identify fetal trisomy we treat 1.049 as the baseline euploid chromosome ratio.

Receiver operating characteristic (ROC) analysis was performed on the complete data set to identify the optimal threshold to separate trisomy 21 spiked samples from the euploid samples using the ratio of Chr21/Chr18 as the predictor. The calculations were performed using the R software package pROC [27, 28]. The optimal threshold was determined to be 1.0672 (Figure 4).

In the tables above σ_{Ratio}, or the standard deviation of calculated chromosome ratio across all replicates, is the most important metric for determining the level of

Table 1. Contrived Trisomy Pregnancy Results

Sim. Aneuploid Fraction	Sample Size	VPdPCR Analysis						Threshold Analysis	
		$\overline{\lambda}_{\text{Chr18}}$	$\overline{\lambda}_{\text{Chr21}}$	Ratio*	ΔRatio	σ_{Ratio}	Accuracy	σ_{Ratio}	Accuracy
7k Haploid Genomic Copies per Well									
0%**	108	0.2666	0.2794	1.048	-0.1%	0.86%	98.1%	1.46%	85.1%
5%	36	0.2658	0.2864	1.078	2.9%	0.83%	97.2%	1.72%	77.8%
10%	36	0.2641	0.2912	1.103	5.4%	0.67%	100%	1.58%	100%
20%	34	0.2632	0.3037	1.154	10.5%	0.86%	100%	1.51%	100%
3.5k Haploid Genomic Copies per Well									
0%**	108	0.1325	0.1390	1.049	0.0%	1.02%	98.1%	1.28%	87.0%
5%	36	0.1330	0.1435	1.080	3.1%	0.95%	88.9%	1.39%	86.1%
10%	36	0.1337	0.1475	1.103	5.4%	0.80%	100%	1.33%	100%
20%	36	0.1327	0.1532	1.154	10.5%	1.10%	100%	1.48%	100%

*$\overline{\lambda_{\text{Chr21}}/\lambda_{\text{Chr18}}}$

**Includes the three euploid only Cell line DNA mixtures from the Table 2: Contrived Euploid Pregnancy Results.

Table 2. Contrived Euploid Pregnancy Results

Sim. Fetal Fraction	Sample Size	VPdPCR Analysis						Threshold Analysis	
		$\overline{\lambda}_{\text{Chr18}}$	$\overline{\lambda}_{\text{Chr21}}$	Ratio*	ΔRatio	σ_{Ratio}	Accuracy	σ_{Ratio}	Accuracy
7k Haploid Genomic Copies per Well									
0%	36	0.2666	0.2794	1.048	-0.1%	0.82%	100%	1.16%	88.8%
5%	36	0.2633	0.2764	1.050	0.1%	1.00%	94.4%	1.13%	88.8%
100%	36	0.2699	0.2823	1.046	-0.3%	0.70%	100%	1.91%	77.8%
3.5k Haploid Genomic Copies per Well									
0%	36	0.1310	0.1376	1.050	0.1%	1.12%	94.4%	1.37%	80.6%
5%	36	0.1317	0.1388	1.054	0.5%	0.72%	100%	1.09%	88.8%
100%	36	0.1347	0.1408	1.045	-0.4%	0.97%	100%	1.36%	91.7%

*$\overline{\lambda_{\text{Chr21}}/\lambda_{\text{Chr18}}}$

quantitative accuracy the assay has achieved. A lower σ_{Ratio} indicates that the assay is more able to precisely identify the true ratio of chromosomes present in the sample, thereby increasing its accuracy in high-precision applications like fetal trisomy testing. Tables 1 and 2 compare σ_{Ratio} when samples were analyzed using VPdPCR (Equation 2) versus the traditional method (Equation 1). The traditional analysis was conducted using sample-specific positive/negative amplitude thresholds for each well by taking the midpoint of the fitted 0-target and 1-target peaks in each channel. Even with this optimized thresholding, the VPdPCR analysis consistently achieved lower σ_{Ratio} on every set of replicates when compared to traditional analysis. This difference was most pronounced for 7k input samples, where VPdPCR cut σ_{Ratio} by more than a factor of 2 in some cases. For 3.5k input samples we expect VPdPCR to provide less of an advantage, as partitioning error is less pronounced at lower input concentrations due to less oversaturation. This theory is reflected in the results, which show a smaller but still consistent improvement from applying VPdPCR in these samples. The differences in ratio distributions from the two analyses are shown visually in Figure 5.

Table 1 show the accuracy of our assay when applying the ROC-optimized threshold to separate euploid from aneuploid samples. 0% simulated fetal fractions samples were called correct if the computed chr21/chr18 ratio fell below the ROC-threshold, and all other samples were called correct of their ratio fell above the threshold. As shown in Figure 5 the ratio distributions for 0% and 5% simulated fetal fractions overlap

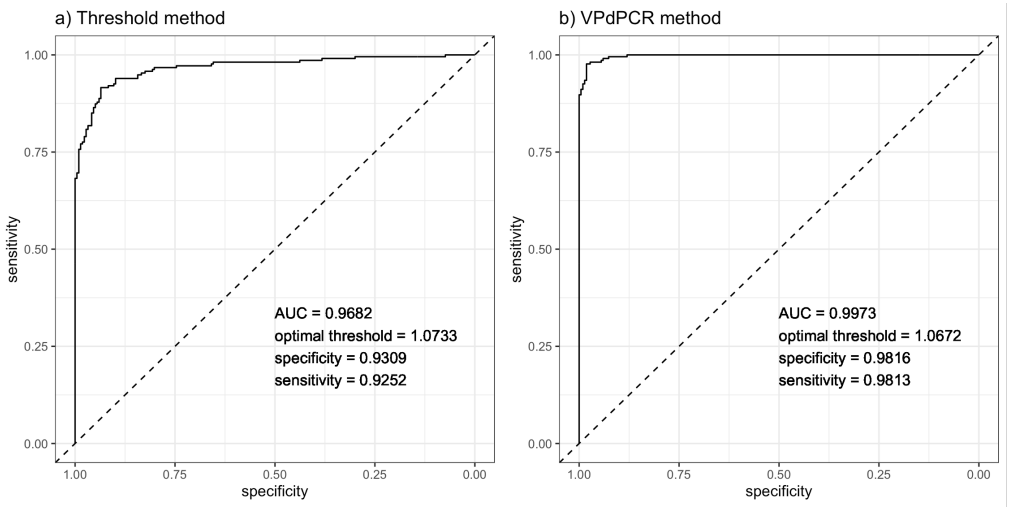

Fig 4. Receiver operating characteristic curve. ROC curves generated using a) the traditional threshold method and b) VPdPCR method. Both ROC curves use the the complete data set including all fetal fractions and both starting concentrations to determine the optimal threshold to separate the pure euploid from trisomy 21 spiked samples.

significantly less when VPdPCR analysis is applied, and this is reflected in the accuracy results; VPdPCR consistently classifies samples with higher accuracy than traditional analysis does. It is worth emphasizing that this data represent a proof of concept for the VPdPCR technique and are not intended to demonstrate clinical viability. This proof of concept demonstrates that VPdPCR has substantial promise for increasing the utility of digital PCR in applications like fetal trisomy screening requiring ultra-high-precision quantitation.

Materials and methods

Novel dPCR Analysis: Multi-Gaussian Fitting

In dPCR analysis the goal is to determine the number of copies per partition of each target, denoted by λ. For multi-channel assays, if we assume targets are independently distributed we can treat each channel independently and thereby compute a separate λ for each channel; this approach is taken for all presented analyses (Figure 3b). The distribution of each target amongst all partitions is dictated by Poisson statistics, which specify that the probability of a partition being negative for a target with concentration λ is simply $p_{neg} = e^{-\lambda}$. In traditional single-target dPCR analysis a single amplitude threshold is drawn to separate positive from negative partitions, and target concentration (in copies per partition) is calculated as $\lambda = -\ln(p_{neg})$ (where p_{neg} is the fraction of partitions below the threshold). If T targets are present at identical concentrations these equations become

$$\begin{aligned} p_{neg} &= e^{-T\lambda} \\ \lambda &= -\frac{1}{T}\ln(p_{neg}) \end{aligned} \tag{1}$$

In real experiments, the precision with which we can determine p_{neg} is limited by the number of negative partitions, with fewer negative partitions leading to higher variance.

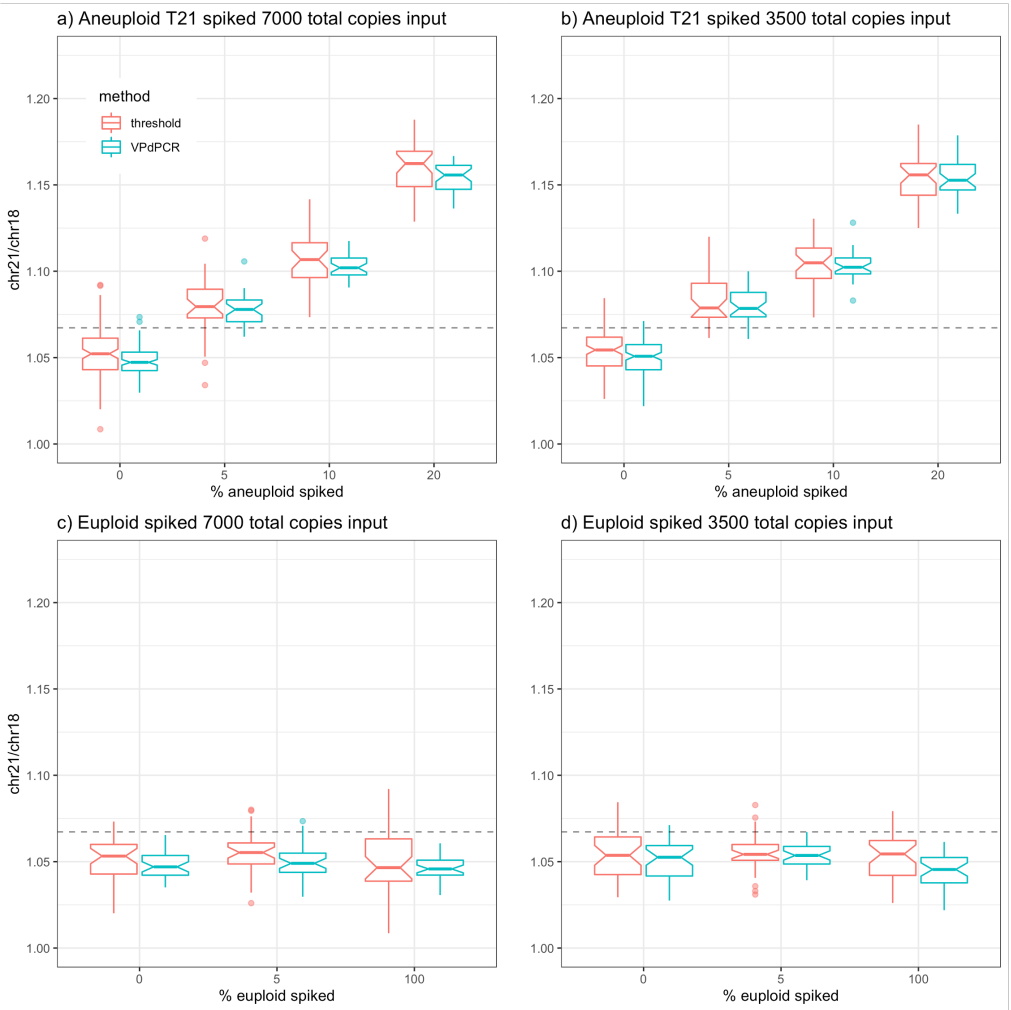

Fig 5. Ratio of Chromosome 21 to Chromosome 18. The ratio Chr21 to Chr18 across a range of euploid and trisomy 21 DNA spiked into euploid background DNA using both the traditional threshold (red) and virtual partition (blue) analysis methods. a) Samples spiked with a T21 cell-line DNA into a euploid background with 7000 haploid genomic equivalents of chromosome 18 per reaction. b) Samples spiked with a T21 cell-line DNA into a euploid background with 3500 haploid genomic equivalents of chromosome 18 per reaction. c) Samples spiked with a euploid cell-line DNA into a different euploid cell-line background with 7000 haploid genomic equivalents of chromosome 18 per reaction. d) Samples spiked with a euploid cell-line DNA into a different euploid cell-line background with 3500 haploid genomic equivalents of chromosome 18 per reaction.
In the box and whiskers plots the center line corresponds to the median, the lower and upper boxes represent the first and third quartiles respectively, and the whiskers extend from the boxes to the smallest and largest values no further from the median than 1.5 times the inter-quartile range. Outlier data beyond the whiskers is represented by individual points and the notch within the boxes approximates the 95% confidence interval of the median. The dashed line is the optimal threshold of 1.0672 as determined by ROC analysis for all samples using the VPdPCR method. Plots were generated with the R software package ggplot2 [29].

This is what leads to the oversaturation effect described earlier. For example, if $T = 10$ and $\lambda = 1$ we get $p_{neg} \approx 5 \times 10^{-5}$, which corresponds to only a single negative partition on average in a 20,000 partition system. This leads to extremely high errors with traditional analysis, necessitating new analysis techniques in such a regime.

To perform more accurate analysis, we need to identify how many of the T targets are present in each partition rather than merely determining whether or not the partition is negative for all targets. This can be done by dividing up the amplitude range of our partitions into bins with indices $t = 0, 1, \ldots, T$. Once bin boundaries are determined, the number of targets present in each partition can be counted by simply determining which boundaries its amplitude falls in between. In principle, on a perfectly consistent system one could run calibration wells and manually draw boundaries between all peaks, then apply those boundaries to sample data. However, on real data, peak locations vary from sample to sample due to a combination of instrument and pipetting variance, leading to poor performance with fixed bin boundaries.

We instead use a more robust method of multi-peak fitting which takes advantage of two observed properties of the system: 1) each peak in the probability density function (PDF) of partition amplitudes can be well approximated by a Gaussian function $\left(G_t(x) = A_t * e^{-(x-\mu_t)^2/2\sigma_t^2}\right)$, and 2) peak amplitudes add linearly, with equal spacing between each subsequent pair of peaks. The fit is based on 5 free parameters: Target Region Copies per droplet (λ, assumed to be the same for all target regions on a chromosome); centers of the 0-target and 1-target bins (μ_0 and μ_1); and widths of the 0-target and 1-target bins (σ_0 and σ_1) The linearity of the system allows us to determine the center and width of all subsequent peaks:

$$\mu_t = \mu_0 + t * (\mu_1 - \mu_0)$$
$$\sigma_t^2 = \sigma_0^2 + t * \left(\sigma_1^2 - \sigma_0^2\right)$$

We can also determine the heights of all peaks based off of 1) Poisson statistics that dictate the probability $P(t)$ of a partition containing t target regions and 2) the fact that the area under a Gaussian curve is equal to $\sqrt{2\pi}\sigma_t A_t$:

$$P(t) = \binom{T}{t} \left(1 - e^{-\lambda}\right)^t \left(e^{-\lambda}\right)^{T-t}$$
$$A_t = \frac{1}{\sqrt{2\pi}\sigma_t} P(t)$$

Once we have calculated μ, σ, and A for every peak we construct a full predicted PDF by adding all of the Gaussian functions together (Figure 6a). The optimal set of $(\lambda, \mu_0, \mu_1, \sigma_0, \sigma_1)$ is determined to be the one which minimizes the RMS error between the full predicted PDF and the observed PDF. This fit is then used to determine $n(t)$, the total number of partitions in each bin. Rather than assigning each one of the N total partitions to a single bin, we divide it between bins based on the relative magnitude of each bin's Gaussian at the partition's amplitude (x_i), improving classification accuracy for higher-order bins in which the tails of neighboring Gaussians start to blend into each other:

$$n(t) = \sum_{i=1}^{N} \frac{G_t(x_i)}{\sum_{t'=1}^{T} G_{t'}(x_i)}$$

Creating Virtual Partitions

Once $n(t)$ has been determined for all bins we use it to calculate target copies. To do this we divide each of our N total partitions into T "virtual" partitions, of which t are

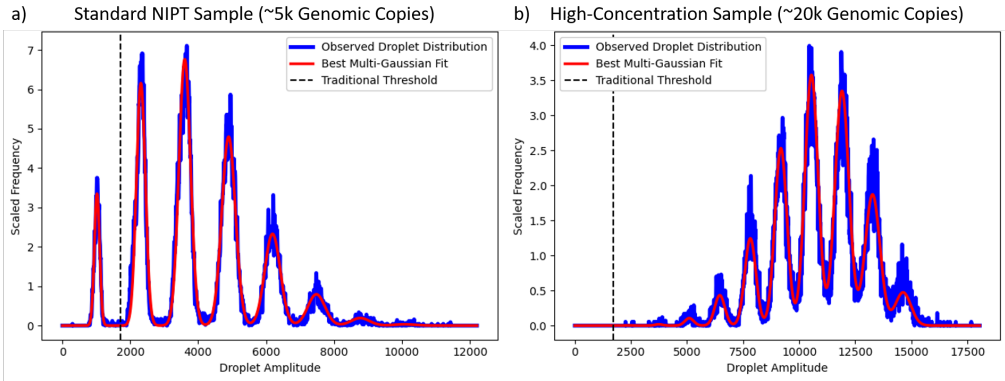

Fig 6. Multi-Gaussian Fitting. Rather than analyzing droplet amplitudes using a traditional positive/negative cutoff threshold (dashed black lines), using VPdPCR we fit the whole amplitude distribution to an extrapolated series of Gaussian functions (red lines). a) Our model matches the observed distribution well in a contrived cfDNA sample with significant peaks up to level 5. b) In a very high-concentration sample there is no 0-target peak, so traditional threshold analysis would fail completely. However, our multi-Gaussian fitting method is still able to perform an appropriate fit and thereby extract target concentration using VPdPCR analysis.

positive and $T - t$ are negative. This effectively transforms our T-target sample with N partitions into a 1-target sample with $T * N$ partitions. We then count the negative virtual partitions and apply our formula from above to get target concentration:

$$\begin{aligned} p_{neg} &= 1 - p_{pos} \\ &= 1 - \frac{1}{N * T} \sum_{t=0}^{T} t * n(t) \\ \lambda &= -\ln\left(1 - \frac{1}{N * T} \sum_{t=0}^{T} t * n(t)\right) \end{aligned} \tag{2}$$

This formulation uses information from all bins 0 through T rather than just bin 0, allowing for accurate analysis at higher concentrations where few partitions are negative for all targets. As shown in Figure 6b this method even works when no droplets are present in the 0-target bin, a regime in which traditional threshold-based analysis breaks down completely.

Theoretical Error Limits

Before experimentally evaluating the VPdPCR assay, we first determined theoretical optimum performance under different levels of multiplexing. For fetal trisomy testing the relevant analysis output is not the absolute number of copies of any one target but rather the ratio between total copies from one chromosome and total copies from another chromosome. The goal is to be able to consistently distinguish between a chromosome ratio of 1 (corresponding to a euploid mother and fetus) and a higher ratio corresponding to a euploid mother and fetus with a trisomy. Detailed statistical analysis by Dube et al [10] allows us to obtain 95% confidence intervals for calculated chromosome ratios given various true input ratios and partition counts. Figure 7a shows these confidence intervals for 20,000 partitions given simulated samples with a euploid fetus and a triploid fetus with a 5% fetal fraction in extracted cfDNA. At 7000 input

genomic copies the intervals significantly overlap, indicating that a traditional assay with 1 target per chromosome cannot consistently identify fetal trisomy at a fetal fraction of 5%. If we increase the number of target regions per chromosome but maintain traditional threshold-based analysis we can effectively increase the input genomic copies without changing the number of partitions. However, as the graph shows, there is no input value for which the intervals are non-overlapping, indicating that no amount of multiplexing can make this task possible with threshold analysis.

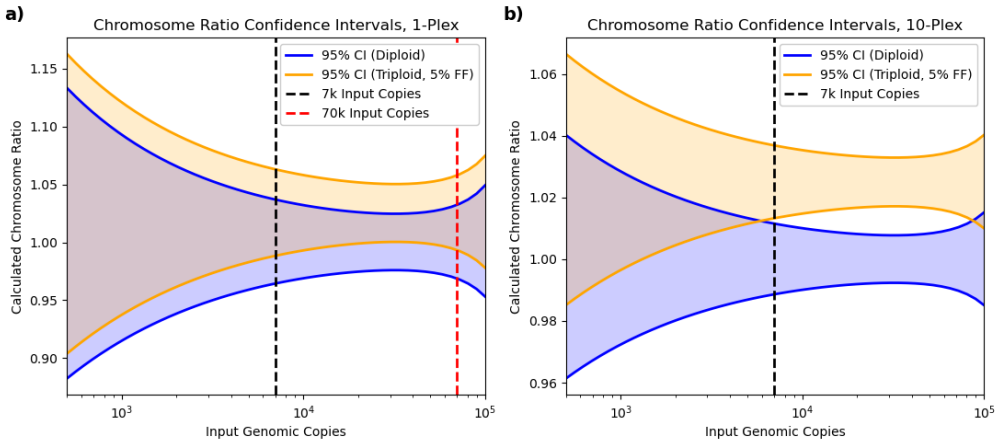

Fig 7. Theoretical Confidence Intervals. To consistently distinguish a diploid from a triploid fetus at a 5% fetal fraction in cfDNA, the two confidence intervals shown should be non-overlapping. a) With a singleplex assay on a machine with 20k physical partitions there is substantial overlap at 7k input copies. Multiplexing with traditional analysis is equivalent to a singpleplex assay with more input copies (red line), but no value of input copies produces non-overlapping intervals. b) Our 10-plex-per-chromosome VPdPCR assay expands the number of effective partitions, creating an input copy region with non-overlapping intervals and making it theoretically viable for fetal aneuploidy testing.

The conventional limitations of traditional threshold analysis changes significantly when we apply multi-Gaussian fitting and VPdPCR analysis to encompass all peaks in multiplexed reactions rather than simple positive/negative classification, which increases both the input copy number and the number of virtual partitions by a factor equal to the number of target regions per chromosome. Figure 7b-d shows the effect of this enhanced multiplexing on chromosome ratio confidence intervals. We found that 10 targets per chromosome is just enough to get theoretically consistent distinction between diploid and triploid samples at 5% fetal fraction and 7k input copies, so we chose that as the target of our assay design.

Digital PCR Assay Design

Multiple TaqMan® PCR assays were designed to amplify conserved regions of Chromosomes 18 and 21 using the Primer3 command line tool [30,31] with GNU Parallel [32] to process designs efficiently on multiple computer cores. We also used primer3 to calculate the binding energies of all pairwise dimers ([monovalent cation] = 50mM, [divalent cation] = 2.5mM, [dNTP] = 0.8 mM, temperature = 60°C), and assays were removed to eliminate favorable oligo-oligo interactions until we reached 10 assays per chromosome. The selected primers and probes were ordered from Integrated DNA Technologies, Inc. (Coralville, IA). Chromosome 18 and 21 TaqMan probes were labeled to be detected in dye channel 1 and channel 2 respectively, and both target

chromosomes' TaqMan probes were double quenched with ZEN™ quenchers. . A 20-plex oligo mix was prepared with all the primers at equal concentration and probes at a significantly lower concentrations. The assay is in development and has not been officially released or approved by the U.S. Food & Drug Administration.

Sample Preparation

Cell line DNA stocks from the Coriell Institute for Medical Research (NA04965 (Trisomy 21), NA12878 (euploid), and NA15453 (euploid)) were sheared with a Covaris® E220 Focused-ultrasonicator (SKU500239) in order to have a mean length of 150 base pairs, simulating the short fragments found in cfDNA. Sheared DNA was processed through the standard singleplex ddPCR workflow using a chromosome 18 target to calculate concentration. To simulate presence of Trisomy 21 stocks were diluted in 1x, low EDTA TE (Gbiosciences 786-151) to a total of 3500 copies/5 μL and 7000 copies/5 μL at 5%, 10% and 20% NA04965 in NA12878. To simulate euploid samples the euploid cell lines were diluted to a total of 3500 copies and 7000 copies/5μl at 0%, 5% and 100% NA15453 in NA12878.

Digital PCR Methods

PCR reactions were set up using the following volumes: 10 μL 2x ddPCR Supermix for probes (no dUTP) (BioRad Laboratories© 186-3024), 5 μL of 20plex oligo mix, and 5 μL of sample. 15 μL of PCR mix was added to each well of 96 well ddPCR plate (BioRad Laboratories 12001925) followed by 5 μL of each sample. Plates were sealed using pierceable foil heat seal (BioRad Laboratories 1814040) and the PX1 plate sealer (BioRad Laboratories 1814000). Plates were vortexed, spun down, and run on the Automated Droplet Generator (BioRad Laboratories 1864101) using Automated Droplet Generation Oil for Probes (BioRad Laboratories 1864110) and the DG32™ Automated Droplet Generator Cartridge (BioRad Laboratories 1864108). After droplet generation was completed, thermocycling was performed on the C1000 Touch with the 96 deepwell module (BioRad Laboratories 1840197). Thermocycling was performed as follows: 1. Enzyme activation (95°C for 10 minutes), 2. 45 Cycles consisting of denaturation (95°C for 20 seconds) followed by combined annealing/extension (58°C for 2 min), 3. Enzyme deactivation (98°C for 10 minutes), and 4. A 4°C infinite hold. Signal detection was performed on the QX200 Droplet Reader (BioRad Laboratories 1864001). Experiment was set to ABS, Target 1 was set to Ch1 Unknown, Target2 was set to Ch2 Unknown and Supermix was set to ddPCR Supermix for probes (no dUTP). Wells were read in columns. Data was exported using the BioRad Laboratories QuantaSoft Version 1.7.4.0917 Software and analyzed using the Python Programming Language version 3.7 (Python Software Foundation, https://www.python.org/).

Conclusion

Digital PCR enables best-in-class rapid quantitative percision for monitoring genomic fragments. However, dPCR on its own, is insufficient for some clinical diagnostic applications, including non-invasive prenatal testing. VPdPCR method not only enhances single-well dPCR multiplexing by a factor of 10 in this demonstration, but also enables dPCR platforms to overcome fundamental limitations to precision by decreasing both sampling and partition error. With newer multi-channel digital platforms, VPdPCR could enable a complete aneuploidy panel for chromosomal abnormalities (Chr21, Chr18, Chr13, X, Y) in a single well. We believe the enhanced precision of VPdPCR could also be useful in a variety of other diagnostic settings, such

as detecting copy number variation of crucial genes to perform liquid biopsies or analyzing low abundance mRNA expression. This range of applications has the potential to make VPdPCR a standard of practice for precision molecular diagnostics.

Supporting information

Table S1. Digital 20plex Processed Data Table Key
File Name: S1_vpdpcr_processed_data.csv

Column Name	Key
run	run number
well_ind	well id numeric 0 indexed
well_name	well location on 96 well plate
sample_name	sample name total DNA_% spiked DNA
partition_ind	partition id numeric 0 indexed
ch1_amp	partition raw fluorescence amplitude data for channel 1
ch2_amp	partition raw fluorescence amplitude data for channel 2
qc_pass	individual partition quality control
well_excluded	wells labeled 1 excluded from analysis

Acknowledgments

The authors would like to thank Jeff Gole and Mimi Wang of ChromaCode, Inc. for making software available to select conserved genomic regions for primer design, Sheila Rosenburg of ChromaCode, Inc for critique and lively statistical discussions and Molly Smith of ChromaCode, Inc for procuring the sheared DNA from UCSD. The following cell lines/DNA samples were obtained from the NIGMS Human Genetic Cell Repository at the Coriell Institute for Medical Research: NA12878DNA, NANA15453DNA and NA04965DNA. Covaris DNA Shearing was conducted at the IGM Genomics Center, University of California, San Diego, La Jolla, CA.

References

1. Whale A S, Huggett J F, Cowen S, et al. Comparison of microfluidic digital PCR and conventional quantitative PCR for measuring copy number variation. Nucleic Acids Res. 2012;40(11):e82
2. Hindson C M, Chevillet J R, Briggs H A, et al. Absolute quantification by droplet digital PCR versus analog real-time PCR. Nat Methods. 2013;10(10):1003-1005.
3. Ricchi M, Bertasio C, Boniotti MB, Vicari N, Russo S, Tilola M, Bellotti MA, Bertasi B. Comparison among the Quantification of Bacterial Pathogens by qPCR, dPCR, and Cultural Methods. Front Microbiol. 2017;8:1174
4. Falzone L, Musso N, Gattuso G, et al. Sensitivity assessment of droplet digital PCR for SARS-CoV-2 detection. Int J Mol Med. 2020;46(3):957-964
5. Huggett JF, Cowen S, Foy CA. Considerations for Digital PCR as an Accurate Molecular Diagnostic Tool. Clin Chem 2015;61(1):79-88.

6. Filetti V, Falzone L, Rapisarda V, et al. Modulation of microRNA expression levels after naturally occurring asbestiform fibers exposure as a diagnostic biomarker of mesothelial neoplastic transformation. Ecotoxicol Environ Saf. 2020;198:110640

7. Salemi R, Falzone L, Madonna G, et al. MMP-9 as a Candidate Marker of Response to BRAF Inhibitors in Melanoma Patients With BRAFV600E Mutation Detected in Circulating-Free DNA. Front Pharmacol. 2018;9:856

8. Battaglia R, Palini S, Vento M E, et al. Identification of extracellular vesicles and characterization of miRNA expression profiles in human blastocoel fluid. Sci Rep. 2019;9(1):84

9. Franke G-N, Maier J, Wildenberger K, Cross M, Giles FJ, Muller MC, Hochhaus A, Niederwieser D, Lange T. Comparison of Real-Time Quantitative PCR and Digital Droplet PCR for BCR-ABL1 Monitoring in Patients with Chronic Myeloid Leukemia. J Mol Diag. 2020;22(1):81-89.

10. Dube S, Qin J, Ramakrishnan R. Mathematical Analysis of Copy Number Variation in a DNA Sample Using Digital PCR on a Nanofluidic Device. PLoS ONE. 2008;3(8):e2876.

11. Zhong Q, Bhattacharya S, Kotsopoulos S, et al. Multiplex digital PCR: breaking the one target per color barrier of quantitative PCR. Lab on a Chip. 2011;11:2167-2174.

12. Taly V, Pekin D, Benhain L, et al. Multiplex Picodroplet Digital PCR to Detect KRAS Mutations in Circulating DNA from the Plasma of Colorectal Cancer Patients. Clin Chem. 2013;59(12):1722-1731

13. Dobnik D, Stebih D, Blejec A, Morisset D, Zel J. Multiplex quantification of four DNA targets in one reaction with Bio-Rad droplet digital PCR system for GMO detection. Sci Rep. 2016;6:35451

14. Alcaide M, Cheung M, Hillman J, et al. Evaluating the quantity, quality and size distribution of cell-free DNA by multiplex droplet digital PCR. Sci Rep. 2020;10:12564

15. Lei S, Gu X, Zhong Q, Duan L, Zhou A. Absolute quantification of Vibrio parahaemolyticus by multiplex droplet digital PCR for simultaneous detection of tlh, tdh and ureR based on single intact cell. Food Control. 2020;114:107207

16. Jacky L, Yurk D, Alvarado J, Belitz P, Fathe K, MacDonald C, Fraser S, Rajagopal A. Robust Multichannel Encoding for Highly Multiplexed Quantitative PCR. Anal Chem. 2021; DOI: 10.1021/acs.analchem.0c04626

17. Allyse M, Minear MA, Berson E, Sridhar S, Rote M, Hung A, Chandrasekharan S. Non-invasive prenatal testing: a review of international implementation and challenges. Int J Womens Health. 2015;7:113-126

18. Benn P, Cuckle H, Pergament E. Non-invasive prenatal testing for aneuploidy: current status and future prospects. Ultrasound Obstet Gynecol. 2013;42:15-33

19. Lo YM, Corbetta N, Chamberlain PF, Rai V, Sargent IL, Redman CW, Wainscoat JS. Presence of fetal DNA in maternal plasma and serum. Lancet. 1997;350(9076):485-7.

20. Samura O. Update on non invasive prenatal testing: a review based on current worldwide research. J Obstet Gynaecol Res 2020;8:1146–1254.

21. Scott FP, Menezes M, Palma-Dias R, Nisbet D, Schluter P, da Silva Costa F, McLennan AC. Factors affecting cell-free DNA fetal fraction and the consequences for test accuracy. J Maternal-Fetal & Neonatal Med 2018;31(14):1865-1872

22. El Khattabi LA, Rouillac-Le Sciellour C, Le Tessier D, et al. Could Digital PCR Be an alternative as a Non-Invasive Prenatal Test for Trisomy 21: A Proof of Concept Study. PLoS One 2016;11(5):e0155009.

23. Tan C, Chen X, Wang F, et al. A multiplex droplet digital PCR assay for non-invasive prenatal testing of fetal aneuploidies. Analyst. 2019;144(7):2239–2247.

24. Haidong W, Zhijie Y, Picchiassi E, Tarquini F, Coata G, You W, Youxiang W, Yu C, Di Renzo GC. Non-invasive prenatal testing of fetal aneuploidies using a new method based on digital droplet PCR and cell free fetal DNA. merRxiv. 2020; DOI 10.1101/2020.12.19.20248553

25. Amar Basu Digital Assays Part I: Partitioning Statistics and Digital PCR SLAS TECHNOLOGY: Translating Life Sciences Innovation. 2017;22(4):369-386. doi:10.1177/2472630317705680

26. Papantoniou N, Bagiokos V, Agiannitopoulos K, Kolialexi A, Destouni A, Tounta G, Kanavakis E, Antsaklis A, Mavrou, A RASSF1A in maternal plasma as a molecular marker of preeclampsia. Prenatal Diagnostics. 2013; 33: 682-687. https://doi.org/10.1002/pd.4093

27. R Core Team R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/

28. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez J-C, Müller M. pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics. 2011; 7:77 doi: 10.1186/1471-2105-12-77

29. Hadley Wickham ggplot2: Elegant Graphics for Data Analysis Springer-Verlag New York. 2016. isbn: 978-3-319-24277-4. url=https://ggplot2.tidyverse.org

30. Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth B, Remm M, Rozen S. Primer3—new capabilities and interfaces. Nucleic Acids Research. 2012; 40(15):e115; DOI 10.1093/nar/gks596

31. Koressaar T, Remm M. Enhancements and modifications of primer design program Primer3. Bioinformatics. 2007; 23(10):1289–1291; 10.1093/bioinformatics/btm091

32. Tange O. GNU Parallel 20200722 ('Privacy Shield'). Zenodo. https://doi.org/10.5281/zenodo.3956817