

## An image-to-answer algorithm for fully automated digital PCR image processing

Zhiqiang Yan<sup>a1</sup>, Haoqing Zhang<sup>a1</sup>, Xinlu Wang<sup>a</sup>, Martina Gaňová<sup>b</sup>, Tomas Lednický<sup>b</sup>, Hanliang Zhu<sup>a</sup>, Xiaocheng Liu<sup>a</sup>, Marie Korabečná<sup>c</sup>, Honglong Chang<sup>a</sup>, Pavel Neužil<sup>a\*</sup>

<sup>a</sup>Northwestern Polytechnical University, School of Mechanical Engineering, Department of Microsystem Engineering, 127 West Youyi Road, Xi'an, Shaanxi, 710072, PR China.

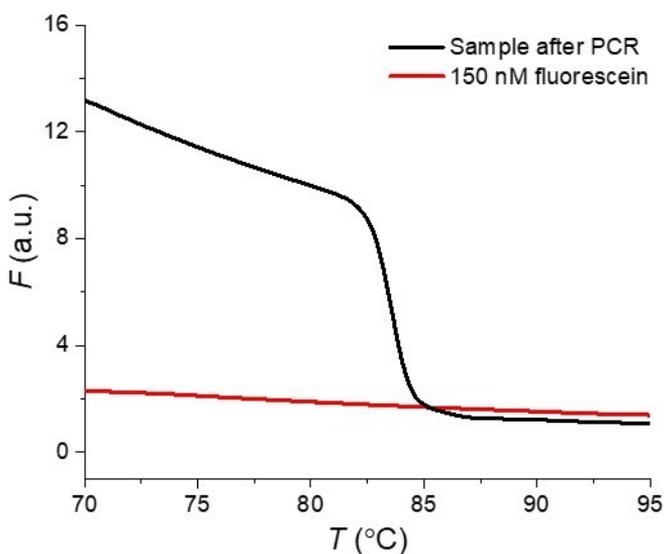
<sup>b</sup>Central European Institute of Technology, Brno University of Technology, Purkyňova 656/123, 612 00, Brno, Czech Republic.

<sup>c</sup>Institute of Biology and Medical Genetics, First Faculty of Medicine, Charles University and General University Hospital in Prague, Albertov 4, 128 00 Prague, Czech Republic.

<sup>1</sup>These two authors contributed equally and both are considered as first authors.

### Supplementary Section A: Temperature sensitivity of the fluorescence

We performed the protocol of melting curve analysis (MCA) using both sample after PCR and 0.15  $\mu\text{M}$  fluorescein. We found that the value of  $F$  from 0.15  $\mu\text{M}$  fluorescein was about 17% of the one from sample after PCR at the same temperature of 70°C (*Fig S 1*). It could be helpful to recognize the negative partitions without affecting the result.



*Fig S 1* MCA protocol showing the  $F$  difference between sample after PCR and fluorescein of 0.15  $\mu\text{M}$ .

Supplementary Section B: The workflow of the proposed algorithm.

The workflow of the proposed algorithm including 3 sections was shown in Fig S 2.

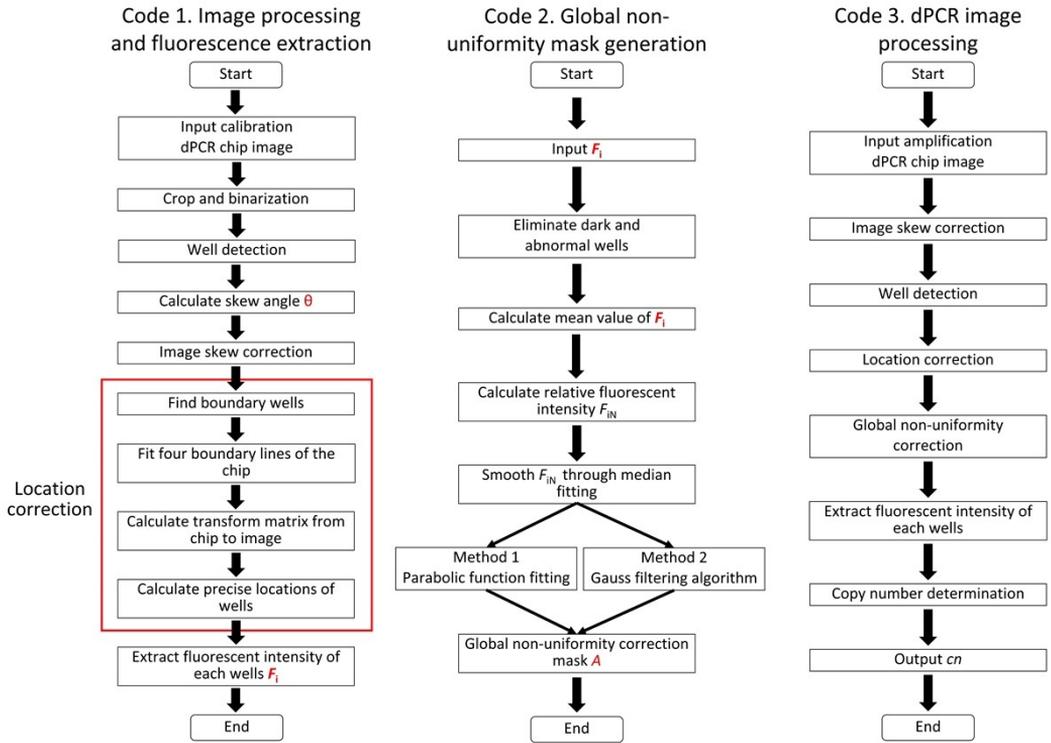


Fig S 2 The proposed algorithm was split into 3 sections. 1<sup>st</sup> section was consisted of image skew correction, well location correction as well as fluorescent amplitude extraction from each well. 2<sup>nd</sup> section was to generate a global light non-uniformity correction mask to remove the influence of non-uniform light distribution on dPCR results. 3<sup>rd</sup> section was to perform the image processing of 1<sup>st</sup> section and apply the light non-uniformity correction mask of 2<sup>nd</sup> section for the dPCR images after thermal cycling, constructing the well number occurrence as a function of fluorescence.

### Supplementary Section C: Details of radial lens distortion correction algorithm.

A radial lens distortion is a common form of image distortion. Here, we take an image (Fig S 3A) of dPCR chip with obvious lens distortion using smartphone (Huawei P40) as an example to demonstrate our method. In the image, the pincushion distortion is radially symmetrical. The center of distortion is also the center of the image.

Firstly, we build a polar coordinate system and calculate polar coordinate of all pixels in the original image as shown in Fig S 3B. After that, corrected coordinates of these pixels are calculated as shown in Fig S 3C. The polar angle  $\theta$  does not change after correction, the polar coordinate  $r_i$  of pixel  $i$  is corrected based on distortion equation:

$$s_i = \frac{r_i}{1 + kr_i^2}, \quad (1)$$

where  $s_i$  is the corrected polar coordinate of pixel  $i$ ,  $k$  is the radial lens distortion correction coefficient. Then, we build a rectangular coordinate system as shown in Fig S 3D. The corrected coordinates of pixels are converted into rectangular coordinates. Then, cubic spline interpolation algorithm is employed to calculate gray value of all pixels in corrected image. Finally, the corrected image is generated based on these pixels.

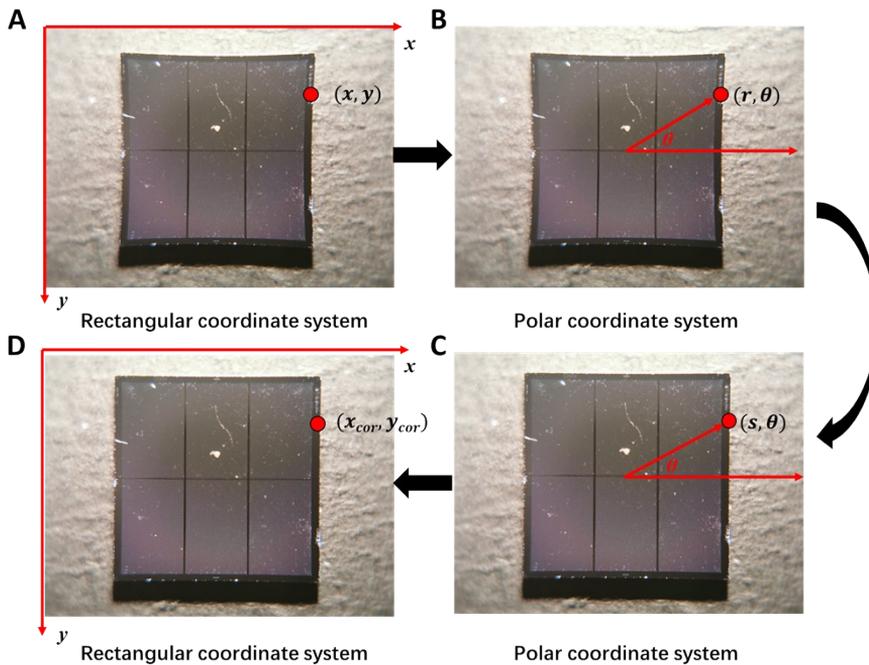


Fig S 3 Schematic illustration of radial lens distortion correction method. (A) Original dPCR image and polar coordinate system built on it. The origin (pole) locates at the center of image, the unit of polar axis is pixel. (B) Corrected image and corrected coordinates of all pixels. The boundaries of dPCR chip are marked with curve. (C) Corrected image and rectangular coordinate system. The origin locates in the upper left corner of the image.

In order to measure the performance of distortion correction and determine the optimal correction coefficient  $k$ , upper boundary of dPCR chip is chosen as a standard. Firstly, 5 uniformly distributed wells located in the upper boundary are manually picked for quadratic polynomial fitting. Quadratic coefficient  $\alpha$  could be obtained after fitting. Then, a series of  $k$  is used for distortion correction. When the absolute value of  $\alpha$  is minimal, its corresponding  $k$  is considered optimal. The boundary curves before and after distortion correction is shown in Fig S 4.

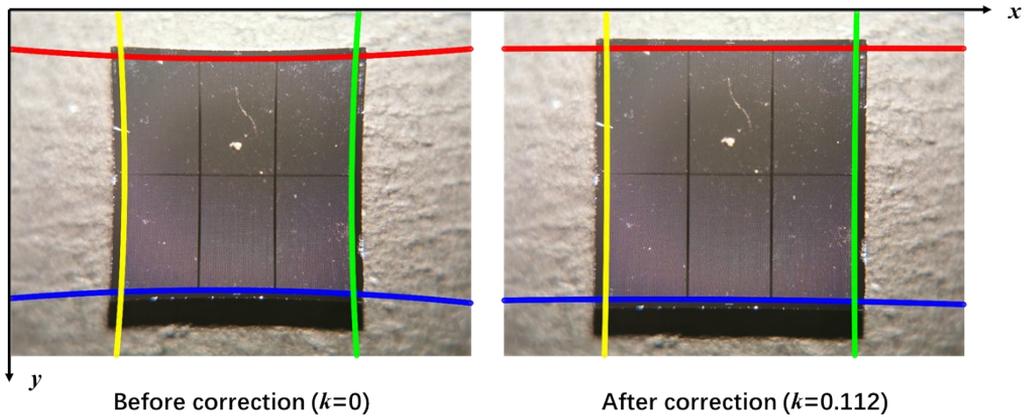


Fig S 4 Boundary curve before and after lens distortion correction. The red, blue, yellow, and green curves are determined by quadratic polynomial fitting of up, down, left and right boundary wells.

Here, we compared quadratic coefficient  $\alpha$  before and after distortion correction as shown in Table S 1. In theory, correction ratio should be very close for all 4 boundaries, however, due to the errors in optical path, they are different from each other. Since we choose upper boundary to measure the degree of distortion, the correction ratio of upper boundary is higher than of other 3 boundaries.

Table S 1 Quadratic coefficient  $\alpha$  before and after correction.

Real angle ( $^{\circ}$ )	Upper boundary	Lower boundary	Left boundary	Right boundary
Before correction	$-1.1747 \times 10^{-5}$	$1.2447 \times 10^{-5}$	$-1.5141 \times 10^{-5}$	$1.2725 \times 10^{-5}$
After correction	$4.8917 \times 10^{-7}$	$3.4444 \times 10^{-6}$	$-3.2275 \times 10^{-6}$	$2.2830 \times 10^{-6}$
Correction ratio	95.84%	72.32%	78.68%	82.06%

## Supplementary Section D: Details of boundary lines fitting algorithm

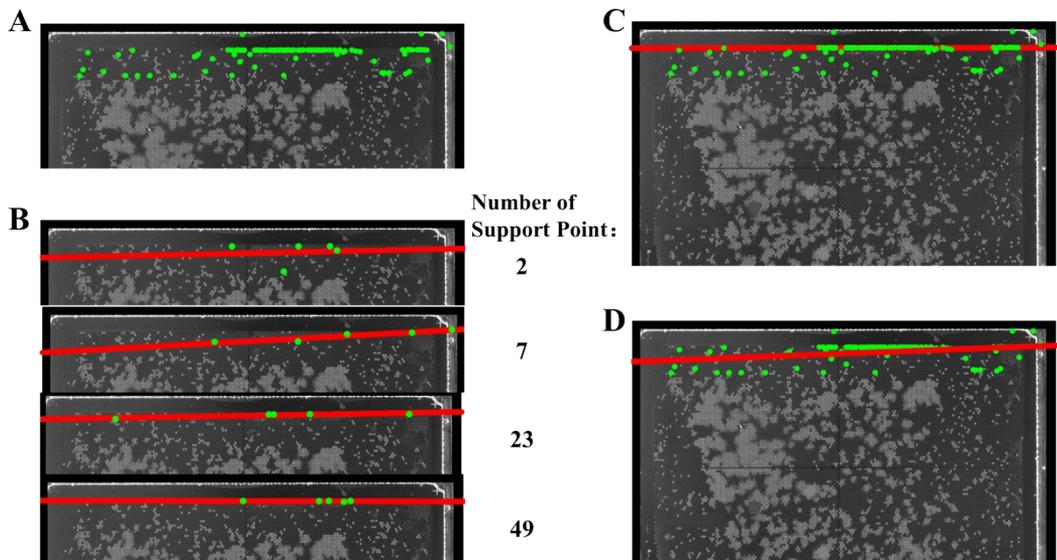


Fig S 5 Schematic diagram of boundary line fitting algorithm. (A) Filtered upper boundary wells. (B) Lines with a different number of support points. (C, D) Fitted upper boundary line using RANSAC method and direct least square method.

To fit boundary lines using boundary wells, including many wrongly detected/filtered ones, we proposed an algorithm based on random sample consensus known as RANSAC method. Here, we take upper boundary as an example to demonstrate it. The filtered upper boundary wells are shown in Fig S 5A.

We chose randomly five wells to fit a line using least square method. After that, the number of wells that were *located* on this line is counted. Such condition is judged by calculating vertical distance between the line and well: if the distance is lower than 5 pixels (which could be adjusted to adapt different cases), this well is regarded located on this line.

Subsequently, the above two steps are repeated 2000 times. The lines with different support point number are shown in Fig S 5B. The corresponding line is considered optimal when the support point number is maximum. Finally, the upper boundary (Fig S 5C) is determined by line fitting of all wells located on this optimal line using the least square method. Compared with direct least square fitting (Fig S 5D), this algorithm could eliminate the effect of a few wrongly detected/filtered wells on boundary location.

## Supplementary Section E: Details of projection transform

2D fluorescence image is a projection of a real dPCR chip. With the location of 4 corner points of the chip, the projection transformation matrix  $T$  between image and chip can be calculated. The wells on dPCR chip are pre-defined and precisely manufactured; therefore, the location of all wells on the chip could also be calculated based on  $T$  as shown in Fig S 6A. This process is also schematically illustrated in Fig S 6B.

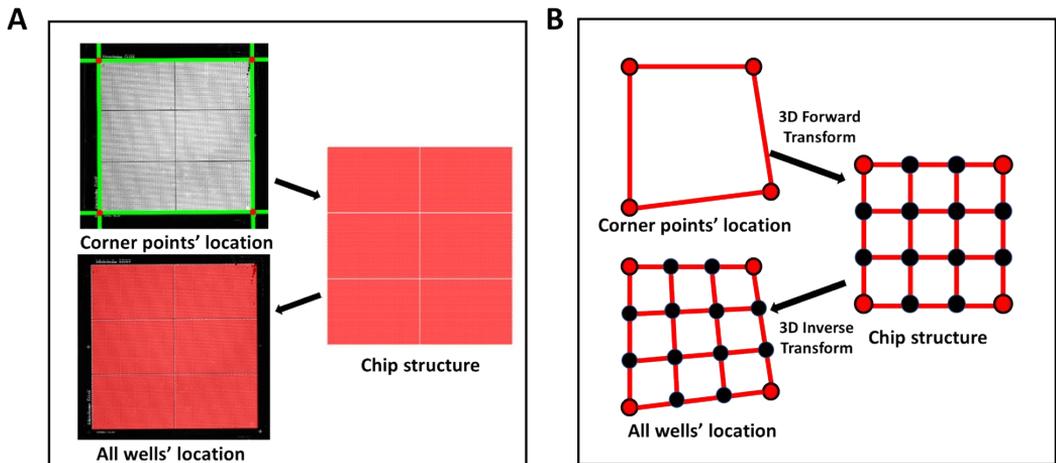


Fig S 6 Schematic illustration of well location correction method. (A) Schematic of location correction method. Four corner points (red) calculated by four boundary lines (green) were used as control points to locate all wells based on pre-defined chip structure. (B) Schematic illustration of 3D projection transform based well location correction method.

Compared with conventional 2D transform method, our method could perform better if the chip's surface is not vertical to the optical path. The 2D and 3D correction methods are compared in Fig S 7. This issue will probably occur if there is a hardware' assembly error, optical instrument's fabrication error or human operation error. The bias between theoretical position and real position increases rapidly with the increase of  $z$ -axis skewed angle ( $\alpha$ ).

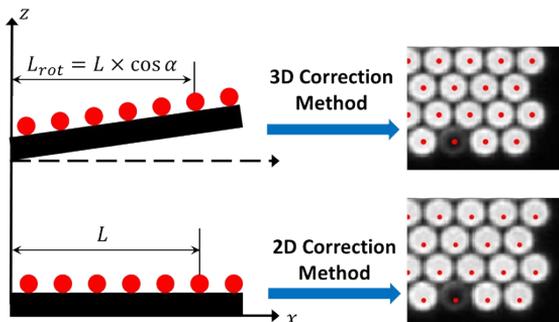


Fig S 7 Comparison of 2D and 3D correction method.

## Supplementary Section F: dPCR images and corresponding histograms using dPCR chips with partition diameter of 50 $\mu\text{m}$

We prepared the master mix with a series  $cn$  values and used it to fill dPCR chips with partition diameter (well) of 50  $\mu\text{m}$  corresponding to the  $\lambda$  values of 0.1, 0.2, 0.4, 0.7, and 0.8. We captured the fluorescent images of the dPCR chips after thermal cycling and performed image processing based on the proposed algorithm and built up the histogram of occurrence as a function of  $F$  value. We determined the number of  $PW$  and  $NW$  based on the histogram and finally used a derived Poisson distribution Eq. (11) to obtain the  $\lambda$  values. The images and histograms were shown in Fig S 8, Fig S 9, and Fig S 10.

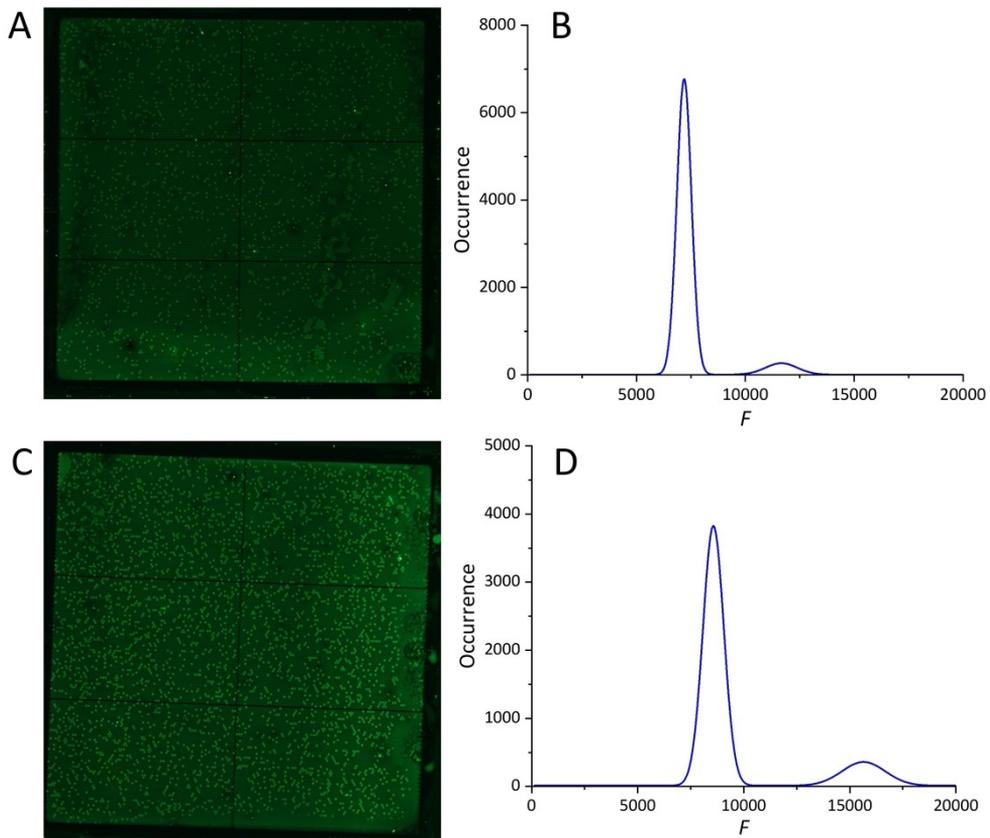


Fig S 8 Fluorescent signal analysis of an amplified dPCR chip with  $\lambda$  values of 0.1 and 0.2. The image (A) and extracted histogram (B) of the dPCR chips after PCR, corresponding to the  $\lambda$  value of 0.1. The image (C) and extracted histogram (D) of the dPCR chips after PCR, corresponding to the  $\lambda$  value of 0.2. We obtained the calculated  $\lambda$  values of 0.105 and 0.173, respectively, showing the relative error of 0.49% and 2.67%, respectively.

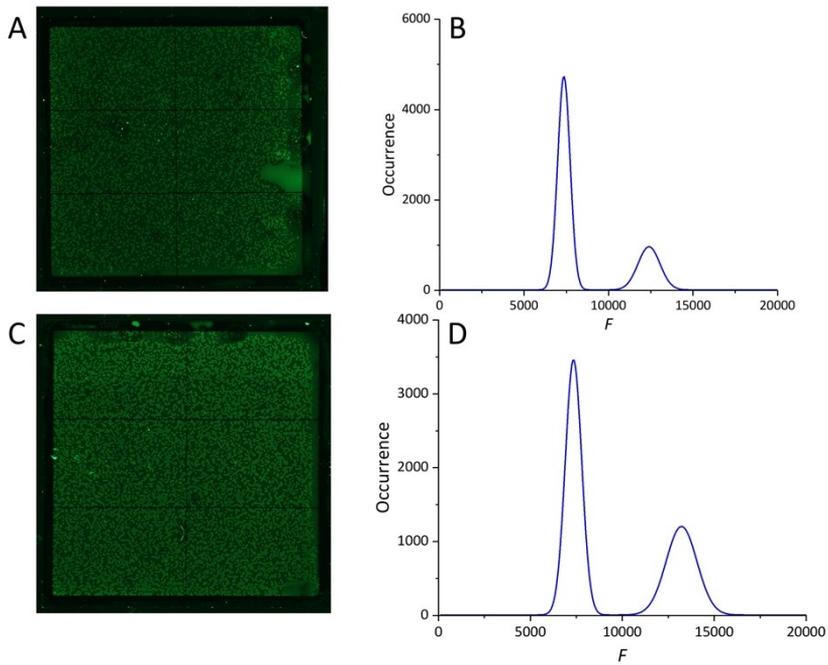


Fig S 9 Fluorescent signal analysis of an amplified dPCR chip with  $\lambda$  values of 0.4 and 0.7. The image (A) and extracted histogram (B) of the dPCR chips after PCR, corresponding to the  $\lambda$  value of 0.4. The image (C) and extracted histogram (D) of the dPCR chips after PCR, corresponding to the  $\lambda$  value of 0.7. We obtained the calculated  $\lambda$  values of 0.347 and 0.564, respectively, showing the relative error of 5.33% and 13.58%, respectively.

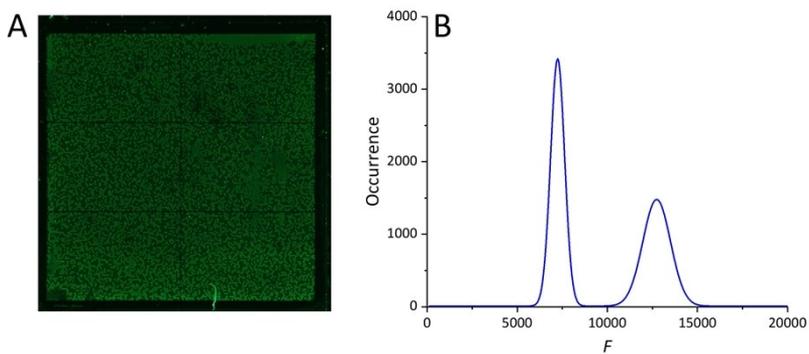


Fig S 10 Fluorescent signal analysis of an amplified dPCR chip with  $\lambda$  value of 0.8. The image (A) and extracted histogram (B) of the dPCR chips after PCR, corresponding to the  $\lambda$  value of 0.8. We obtained the calculated  $\lambda$  values of 0.676, showing the relative error of 12.41%.

Supplementary Section G: dPCR images and corresponding histograms using dPCR chips with partition diameter of 20  $\mu\text{m}$

We filled dPCR chips with partition diameter (well) of 20  $\mu\text{m}$  corresponding to the  $\lambda$  value of 0.1. We then captured three blocks of the fluorescent images of the dPCR chips after thermal cycling and performed image processing based on the proposed algorithm and built up the histogram of occurrence as a function of  $F$  value. We determined the number of PW and NW based on the histogram and finally used a derived Poisson distribution Eq. (11) to obtain the  $\lambda$  values. The images and histograms were shown in Fig S 11.

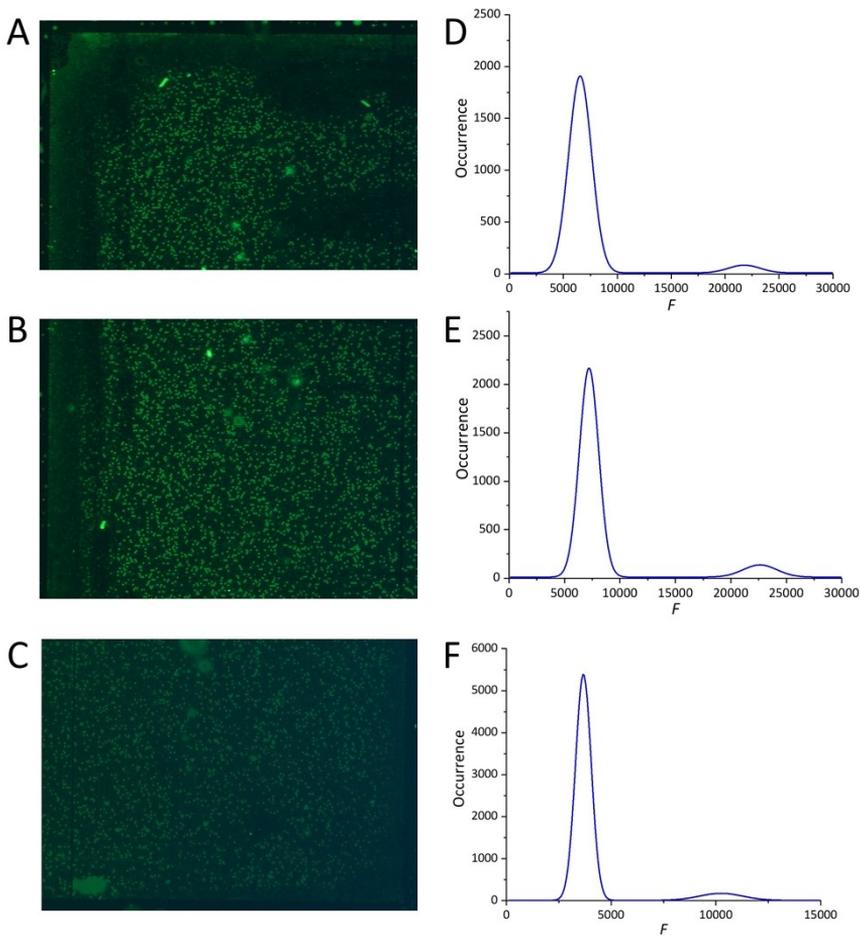
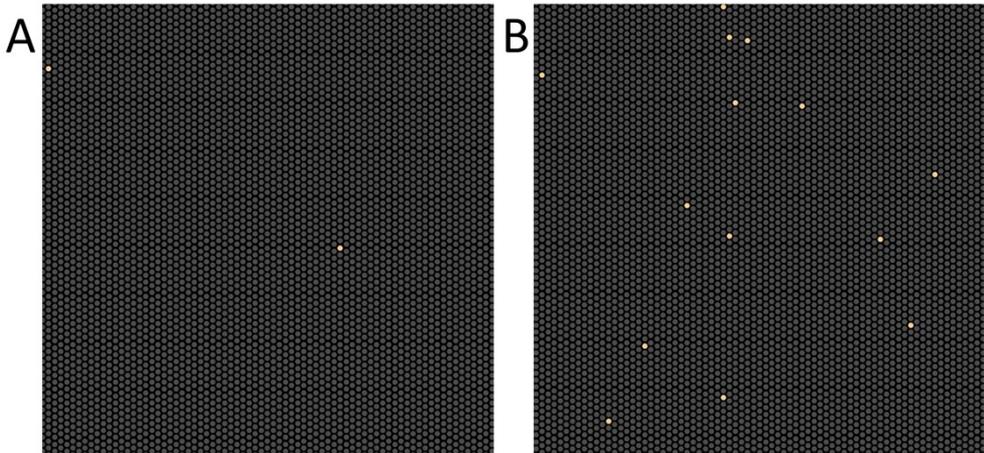


Fig S 11 Fluorescent signal analysis of 3 blocks of a dPCR chip with partition diameter of 20  $\mu\text{m}$ , corresponding to a  $\lambda$  value of 0.1. The images (A, B, C) and extracted histograms (D, E, F) of the dPCR chips demonstrated the calculated  $\lambda$  values of 0.06, 0.09 and 0.11, respectively.

### Supplementary Section H: dPCR images of QuantStudio 3D dPCR chips

We used images from commercial dPCR chips (ThermoFisher Quantstudio 3D) after thermal cycling to validate the algorithm, corresponding to the original  $cn$  values of 2 and 20 (Fig S 12).



*Fig S 12 Fluorescent images from the commercial dPCR chips, corresponding to  $cn$  values of 2 (A) and 20 (B).*