Technical Paper

Analysis Method for Digital PCR Data on Two Dimensional Orthogonal Coordinate \((x, y)\) by Converting to Two Dimensional Polar Coordinate \((r, \theta)\) Using EXCEL Macro

Takashi FUJII*, Mikio KAWAHARA, Akihiro TSUYADA, Manabu SUGIYAMA, Masakazu AKAHORI

Riken Genesis Co., Ltd. 141-0032 Osaka 1-2-2, Shinagawa-ku, Tokyo Japan

*e-mail: takashi.fujii@rikengenesis.jp

(Received: September 29, 2017; Accepted for publication: March 14, 2019; Online publication: April 18, 2019)

An absolute quantitative analysis method has been recently developed as a third generation polymerase chain reaction method “PCR” for fractionated DNA. The method is designed to determine the number of DNA molecules in target DNA samples by counting the number of PCR products obtained from fractionated DNA. We applied EXCEL Macro to perform the conversion of two dimensional orthogonal coordinate \((x, y)\) fluorescent signal plot data obtained by digital PCR device to two dimensional polar coordinate \((r, \theta)\) fluorescent signal plot data, followed by analyzing the angle \((\theta)\) histogram of plot data without overlapping of plot data occurring by two dimensional orthogonal coordinate \((x,y)\) histogram. The analysis made it possible to identify gene mutation and count the number of DNA molecules with mutation faster and easier.

Key words: Digital PCR, DNA, Polar coordinate, EXCEL

1 INTRODUCTION

It has been demonstrated that data from digital PCR analysis [1–4], in which wild type and mutant type are labeled with a probe/Hexachloro Fluorescein (HEX) and a probe/Fluorescein Amidite (FAM), respectively, generate almost 15,000–20,000 fluorescent signals, the results of which are recorded in the two dimensional orthogonal coordinate \((x, y)\) to determine the extent of target gene amplification [5]. It is possible to detect the presence of gene mutation and the rate of mutation with high sensitivity in an extremely small amount of target sample by plotting the intensity of mutant type (Mt) with HEX fluorescent signal in the \(x\)-axis and wild type (Wt) with FAM fluorescent signal in the \(y\)-axis in two dimensional orthogonal coordinate \((x, y)\) (Figure 1). It is, however, difficult to show the distribution status in two dimensional orthogonal coordinate \((x, y)\) histogram without overlapping of the plots between no DNA template control (NTC) region and Mt region or NTC region and Wt region, resulting in the need for two separate histograms created from \(x\) (HEX) axis direction and \(y\) (FAM) axis direction. In this study, it is possible to observe all fluorescent signal plots in angle \(\theta\) histogram without overlapping, where the histogram was created based on the value of angle \(\theta\) of the converted two dimensional polar coordinate \((r, \theta)\) [6], which was prepared by converting from the data in two dimensional orthogonal coordinate \((x, y)\) by EXCEL. The origin of the converted two dimensional polar coordinate \((r, \theta)\) was moved to the gravity point of the triangle consisting of each center of Mt region, Wt region and NTC region. Therefore, it is possible to determine gene mutation easily and accurately as well as determine mutation rate by applying the polar coordinate approach.

2 METHOD AND RESULTS

2.1 Digital PCR data in two dimensional orthogonal coordinate \((x, y)\)

It has been demonstrated that the extent of target gene amplification could be monitored by plotting the intensity of fluorescence signals of Wt/HEX in the \(x\)-axis and those in Mt/FAM in the \(y\)-axis in two dimensional orthogonal coordinate \((x, y)\). The occurrence frequency was only determined by histogram. It is, however, difficult to count the accurate number of Mt/FAM fluorescent signals in \(x\) axis histogram, because of overlapping signals of NTC with those of Mt/FAM (Figure 2). Similarly, it was also difficult to count the accurate number of Wt/HEX fluorescent signals in \(y\) axis histogram, because of overlapping signals of NTC with those of Wt/HEX (Figure 3). Therefore, the distribution condition of Mt/FAM and Wt/HEX fluorescent signals by one histogram is difficult to observe in two dimensional orthogonal coordinate \((x, y)\) without any overlapping signals.

2.2 Conversion from two dimensional orthogonal coordinate \((x, y)\) to two dimensional polar coordinate \((r, \theta)\)

The origin \((x = 0, y = 0)\) in two dimensional orthogonal coor-
Fluorescent signal plots of dPCR by two dimensional orthogonal coordinates \((x, y)\). Sample is ALK wild type/resistant mutation type (L1196M) DNA. A; Mutation Type region (Mt region), B; No Template Control region (NTC region), C; Wild Type region (Wt region). HEX (Hexachloro Fluorescin); a probe for Wild type DNA, FAM (Fluorescein Amidite); a probe for Mutation type DNA. HEX fluorescent signals are plotted in \(x\) axis and FAM fluorescent signals are plotted in \(y\) axis.

Figure 2. Histogram of two dimensional orthogonal coordinate (HEX axis). Sample is ALK wild type/resistant mutation type (L1196M) DNA. In histogram of \(x\) (HEX) axis, A (Mt region) is overlapped with B (NTC region), therefore HEX fluorescent signal plot numbers of A (Mt region) can’t be counted accurately. Two separate histograms of \(x\) (HEX) axis and \(y\) (FAM) axis are needed to calculate the fluorescent signal plot numbers of A (Mt region) and C (Wt region), as same as the caption in Figure 2.

Figure 3. Histogram of two dimensional orthogonal coordinate (FAM axis). Sample is ALK wild type/resistant mutation type (L1196M) DNA. In histogram of \(y\) (FAM) axis, C (Wt region) is overlapped with B (NTC region) like the case of Figure 2, therefore FAM fluorescent signal plot numbers of C (Wt region) can’t be counted accurately. Two separate histograms of \(x\) (HEX) axis and \(y\) (FAM) axis are needed to calculate the signal plot numbers of A (Mt region) and C (Wt region), as same as the caption in Figure 2.

The coordinate \((x, y)\) was moved to the gravity of A (Mt region), B (NTC region) and C (Wt region), and then, two dimensional orthogonal coordinate \((x, y)\) were converted to two dimensional polar coordinate \((r, \theta)\) (Figure 4). In the two dimensional polar coordinate \((r, \theta)\), all plots are observed by \((\theta)\) histogram without any overlapping (Figure 5). The gravity was obtained by cluster analysis method of three elements \((k = 3)\), such as A (Mt region), B (NTC region) and C (Wt region). The gravity calculation was executed automatically by EXCEL Macro [7] for \(k\)-means clustering method [8–10], a typical dividing-optimization method. The explanation for conversion equation from two dimensional orthogonal coordinate \((x, y)\) to two dimensional polar coordinate \((r, \theta)\) is as follows:

\[
\begin{align*}
    r &= \sqrt{x^2 + y^2} \\
    \theta &= \text{DEGREES} \left( \text{ATAN} \left( \frac{y}{x} \right) \right)
\end{align*}
\]
2.3 Determination method for the identification of gene mutation and the rates by two dimensional polar coordinate \((r, \theta)\)

After the coordinate converting, the number of fluorescent signal plots in the A (Mt region), B (NTC region) and C (Wt region) are calculated automatically using the normal distribution probability density function “NORMDIST” in EXCEL at each of 90%, 95% and 99%, respectively. “center” in table is expressing each centroid coordinate \((x, y)\) of A, B and C region, respectively. Inside, middle and outside elliptical graphs are expressing the areas at each confidence interval 90%, 95% and 99%, respectively.

2.4 EXCEL Macro Processing Procedure

It is possible to judge promptly and easily whether target gene mutations are present by visualizing genetic analysis fluorescent signal plots data through the following processing procedure of EXCEL VBA MACROS analysis except that the Steps 1) and 9) were input and evaluated manually.

[Operating Procedure]

1) Data uptake two dimensional orthogonal coordinate \((x, y)\)
2) Determine centroids for corresponding A (Mt region), B (NTC region) and C (Wt region) by k-means clustering method
3) Determine the gravity point of triangle formed by the centroids for A (Mt region), B (NTC region) and C (Wt region)
4) Parallel move of the origin of two dimensional orthogonal coordinate \((x, y)\) to the gravity point
5) Conversion to two dimensional polar coordinate \((r, \theta)\)*
   \[ r = \sqrt{x^2 + y^2}, \quad \theta = \text{DEGREES} (\text{ATAN} (Y/X)) \]
6) Histogram preparation based on \((\theta)\) degrees
7) Calculate the number of fluorescent signal plots distributed in A (Mt region), B (NTC region) and C (Wt region) within 90%, 95% and 99% confidence interval by NORMDIST function
8) Re-conversion** to two dimensional orthogonal coordinate \((x, y)\) and elliptical graph of 90%, 95% and 99% confidence interval display

** \(x = r \times \cos (\text{RADIANS} (\theta)), \quad y = r \times \sin (\text{RADIANS} (\theta))\)

9) Determine gene mutation and calculate gene mutation rates within 90%, 95% and 99% confidence interval, respectively.
DISCUSSION

In quantitative analysis from histograms in HEX (x axis) and FAM (y axis), it is difficult to show each distribution state of the fluorescent signal plots in A (Mt region), B (NTC region) and C (Wt region) at the same time, due to overlapping of fluorescent signal plots in A (Mt region) with those in B (NTC region) or C (Wt region) with B (NTC region) so far. Therefore, the origin \((x = 0, y = 0)\) was moved to the gravity point inside of the triangle consisting of A (Mt region), B (NTC region) and C (Wt region), and the two dimensional orthogonal coordinate \((x, y)\) were converted to the two dimensional polar coordinate \((r, \theta)\). It is confirmed that this approach made it possible to show separately all fluorescent signal plots of the three regions such as A (Mt region), B (NTC region) and C (Wt region) in histogram at one time without any overlapping and easy to count the number of fluorescent signal plots in A (Mt region) and C (Wt region). Furthermore, we consider that employing an automated processing from coordinate conversion to graphic visualization of results by EXCEL Macro is useful for identifying mutation occurrence and mutation rate promptly. For the future, we are hoping to develop a more convenient genetic analysis method by processing more samples using the method described here.

REFERENCES

[1] X. Shi, C. Tang, W. Wang, D. Zhou, Z. Lu, Electrophoresis, 31, 528 (2010). PMID:20119960, DOI:10.1002/elps.200900362
[2] E. A. Ottesen, J. W. Hong, S. R. Quake, J. R. Leadbetter, Science, 314, 1464 (2006). PMID:17138901, DOI:10.1126/science.1131370
[3] B. Vogelstein, K. W. Kinzler, Proc. Natl. Acad. Sci. USA, 96, 9236 (1999). PMID:10430926, DOI:10.1073/pnas.96.16.9236
[4] S. Dube, J. Qin, R. Ramakrishnan, PLoS One, 3, e2876 (2008). PMID:18682853, DOI:10.1371/journal.pone.0002876
[5] B. J. Hindson, K. D. Ness, D. A. Masquelier, P. Belgrader, N. J. Heredia, A. J. Makarewicz, I. J. Bright, M. Y. Lucero, A. L. Hiddessen, T. C. Legler, T. K. Kitano, M. R. Hodel, J. F. Petersen, P. W. Wyatt, E. R. Steenblock, P. H. Shah, L. J. Bousse, C. B. Trup, J. C. Mellen, D. K. Wittmann, N. G. Erndt, T. H. Cauley, R. T. Koehler, A. P. So, S. Dube, K. A. Rose, L. Montesclaros, S. Wang, D. P. Stumbo, S. P. Hodges, S. Romine, F. P. Milanovich, H. E. White, J. F. Regan, G. A. Karlin-Neumann, C. M. Hindson, S. Saxonov, B. W. Colston, Anal. Chem., 83, 8604 (2011). PMID:22035192, DOI:10.1021/ac202028g
[6] Kyoritsu Sugakukosiki, Kyoritsu Shuppan, pp. 111 (1975)
[7] S. Neilson, “k-Means Cluster Analysis in Microsoft Excel” (2011). http://www.neilson.co.za/k-means-cluster-analysis-in-microsoft-excel/
[8] H. Steinhaus, (French). Bull. Acad. Polon. Sci., 4, 801 (1957).
[9] E. W. Forgy, Biometrics, 21, 768 (1965).
[10] J. B. MacQueen, “Proceedings of 5-th Berkeley Symposium on Mathematical Statistics and Probability”, Berkeley, University of California Press, 1, 281–297 (1967).