Bull Yamaguchi Med School 56(3-4):43-48, 2009

The Development of a Variable Threshold Analysis Method to Identify Clones Linked the Clinicopathological Features in an Array-Based CGH－Its Application to Colorectal Cancer

Motonao Nakao,[1] *Tetuji Uchiyama,*[2] *Jun Adachi,*[2] *Tomoko Furuya,*[1] *Atsunori Oga,*[1] *Shigeto Kawauchi*[1] and *Kohsuke Sasaki*[1]

[1] Department of Pathology, Yamaguchi University Graduate School of Medicine, 1-1-1 Minami-Kogushi, Ube, Yamaguchi 755-8505, Japan
[2] Department of Surgery, Iwakuni Medical Center, Iwakuni, Yamaguchi 740-0021, Japan
(Received November 13, 2009, accepted December 21, 2009)

Abstract In an array-based CGH, the analytical results of an evaluation of an abnormal number of copies are substantially affected by the designation of a cut off value for the ratio of the fluorescence intensity. In this study, we developed an array-based CGH analysis method that enables the designation of an optimal cut off value adapted to the analysis. With this method, the threshold of each BAC clone fluctuates automatically depending on the number of samples in order to enable the identification of a BAC clone correlated with the clinicopathological parameters. In other words, a convenient threshold is set so that the significant difference in a kai square test between two groups classified according to a certain clinicopathological parameter is maximized. Moreover, its usefulness has been verified based on 74 cases of colon cancer. When 74 cases of colon cancer were divided into 2 groups -stage I/II and stage III/IV- there was a significant difference (P<0.001) between the two groups only in the gain of the number of clones in region 8q24.3 when testing only with a threshold of ±0.25. Using this method of analysis, the gain of the number of clones copied in region 7q11.23 and the loss of the number of clones copied in regions 20p12.1, 5q33.3, 17p12, and 9q33.1 were also observed in significant levels (p<0.001) in stage III/IV of colon cancer. Meanwhile, the gain of the number of clones copied in region 1q44 and the loss of the number of clones copied in regions 3q13.33, 14q12, 5p15.1, and 14q32.11 correlated with stage I/II colon cancer. The use of this method enabled the efficient identification of clones that were correlated with a specific parameter.

Key words: array based CGH, threshold

Introduction

An array-based CGH has been used most frequently for detecting an abnormal number of DNA copies. There have been various discussions regarding the designation of the threshold of the ratio of fluorescence intensity for array-based CGH, which relies on the determination of the number of DNA copies.[1] A fixed value with a two-bottom log ratio of ±0.2 to 0.5 is used in normal analyses.[2,3] In addition, most articles have used ±0.25 after considering the variations in the fluorescence ratio after normalization. This has been set with the consideration that a threshold of ±0.25 is exceeded when the disomy of a cell in a steady state becomes trisomy or monosomy, and conversely, the threshold of ±0.25 is not ex-

ceeded with only experimental error. Having a constant cut off value makes determination easy. However, the nuclear DNA content, which is the background for the ratio of the increase or decrease of each chromosome and each chromosomal region, is not always diploid. Therefore, it is not necessarily desirable to apply an identical threshold to all chromosomes. In fact, in an array-based CGH of cancer cells, a small variation that does not reach ±0.25 is not uncommon for reasons other than experimental error. Therefore, it is very unreasonable to set a unique threshold in order to simply binarize a gain or loss when analyzing the data of an array CGH, and we believe that it is proper to determine thresholds according to actual measured data.

We have developed a new aCGH analysis method in which the threshold is set freely so as to vary for the detection of genomic aberrations that are correlated with a parameter that is required in clinical practice, and we have verified the usefulness thereof in cases of colon cancer by targeting each stage.

Methods

Method of setting a threshold

The threshold is not simply changed, but rather, the ratio of the fluorescence intensity of a number corresponds to the number of cases (herein, 74). That is, because all of the cases can be classified by examining the threshold corresponding to the number of cases, the ratio of the fluorescence intensity is used as a threshold as is. Accordingly, the calculated amount is 74 x 4,030 clones. The ratio of the fluorescence intensity has been calculated to four decimal places. A threshold of 74 was attempted per 4,030 clones, and this was programmed to save the threshold and the matrix data when the p-value of the kai square test reached its lowest point.

Verification in cases of colon cancer

A total of 74 cases of colon cancer were examined (Table 1). These cases were divided into two groups for convenience: stages I/II (31 cases) and stage III/IV (42 cases). For these cases of colon cancer, we conducted an array-based CGH using a previously reported procedure. Specifically, DNA (500 ng) extracted from a cancer cell that had been selectively isolated from a cancer tissue with tissue microdissection was labeled with Cy 3 (Perkin Elmer, Wellesley, MA) and control DAN with Cy 5 (Perkin Elmer), and for the array (Macrogen, Korea), in which 4,030 BAC clones were spotted as duplicates in the presence of Cot-1 (50 mg, Gibco BRL, Gaithersburg, MD), hybridization was conducted for 72 hours at 37 ℃. A GenePix 4000A scanner (Axon Instruments, Union City, CA) was used for reading the fluorescence signals. The ratio of the fluorescence intensity of Cy3/Cy5 was recorded with a two-bottom log.

Table 1 Clinical data of 74 colorectal cancer

Sex	male	40
	female	34
location	Cecum, Ascending colon Transverse colon	22
	Descending colon, Sigmoid colon	21
	Rectosigmoid, Rectum, Proctos	31
Stage	I or II	32
	III	28
	IV	14

As described above, a threshold of 74 per 4,030 clones was set, and the threshold and the matrix data at the time when the p-value of the kai square test reached its lowest point (the total number of cases at Stages I & II exceeding the threshold and data, and the total number of cases at Stages III & IV exceeding the threshold and data) were saved. In this study, considering the Bonferroni correction, we used the number of clones that were most likely to be significant when $p<0.001$.

Results

For the conventional method of analysis, the threshold was set at ±0.25, and the kai square test was conducted on two groups low-stage and high-stage colon cancer with 0.25 or higher designated as a gain and -0.25 or lower designated as a loss. In the results,

Table 2 BAC clones with a significant difference using X2 test between StageI&II and Stage III&IV using 0.25 threshold

Location	BAC start	plate no.	Gene contained in the clone	P value	Stage I&II	Stage III&IV
8q24.3	146118993	2748	ZNF16, Tmp21-II, LOC441383, LOC286103,	6.94E-05	1 / 31	19 / 42

Table 3 BAC clones with a significant difference using X2 test between StageI&II and StageIII&IV using variable threshold

Location	BAC start	BAC plate no.	Gene contained in the clone	P value	Thereshold	Stage I&II	Stage III&IV
8q24.3	146118993	2748	ZNF16, Tmp21-II, LOC441383, LOC286103,	6.94E-05	0.2483	1 / 31	19 / 42
3q13.33	122354108	1102	STXBP5L,	7.44E-05	−0.15	19 / 31	40 / 41
14q12	30885368	508	C14orf125, LOC400201, C14orf126,	0.000226	−0.3121	22 / 31	41 / 41
20p12.1	16603508	1509	RPL7AL3, SNRPB2, OTOR,	0.000249	−0.1182	30 / 31	23 / 39
7q11.23	75180167	5617	POR, TMPIT, DUSP24,	0.000356	0.178	4 / 30	22 / 40
5q33.3	158772473	2853		0.000483	−0.1388	24 / 28	17 / 39
1q44	243153563	4684	FLJ32001, CGI-49,	0.00056	0.289	8 / 31	0 / 41
5p15.1	18102403	4466		0.000648	−0.2476	23 / 31	40 / 40
17p12	11944544	2859	MAP2K4,	0.000713	−0.2667	28 / 31	20 / 38
9q33.1	1.17E+08	4289		0.000721	−0.1524	27 / 31	20 / 41
14q32.11	89261806	1039	CHORDC2P, C14orf143,	0.000829	−0.329	20 / 31	39 / 41

the clones with P<0.001 are shown in Table 2. When the threshold was fixed, only one clone corresponded to P<0.001 (8q24.3, clone 2748). At Stage I/II, 1 of 31 cases (3.2%) had an increased number of DNA copies of this clone, but an increase was observed in 19 of 42 cases (45.2%) at Stage III/IV.

On the other hand, without setting a constant threshold, a threshold of 74 (corresponding to the number of cases) was set for all of the clones (4,030 clones), the frequency of the occurrence of abnormal copies between Stage I/II and III/IV became the maximum, and 11 clones with P<0.001 and an absolute value of the threshold corresponding to at least 0.1 were identified (Table 3). With regard to the results of the array-based CGH of 74 cases of colon cancer, by varying the threshold, the detection sensitivity for the same significant difference became at least 10 times higher. For clone 2748 (8q24.3), cases with a threshold of >0.2483 included 1 in 31 cases at Stage I/II and 19 in 42 cases at Stage III/IV, which is a result that is similar to that with a fixed threshold. On the other hand, for clone 1102 (3q13.33), which was not detected via the threshold fixation method, the cases that corresponded to a threshold of >-0.15 in which the p value between the two groups was minimal included 19 in 31 cases at Stage I/II and 40 in 41 cases at Stage III/IV. From a different perspective, the cases with a threshold of <-0.15 included 12 in 31 cases at Stage I/II and 1 in 41 cases at Stage III/IV. The number of copies gained for clone 5617 (7q11.23) and the number of copies lost for clone 1509 (20p12.1), clone 2853 (5q33.3), clone 2859 (17p12), and clone 4289 (9q33.1) appeared to be similar in Stage III/IV of colon cancer. On the other hand, the number of copies

gained for clone 4684 (1q44) and the number of copies lost for clone 1102 (3q13.33), clone 508 (14q12), clone 4466 (5p15.1), and clone 1039 (14q32.11) were similar at Stage I/II of colon cancer.

Discussion

aCGH is a powerful tool for making a comprehensive analysis of DNA copy number aberrations in the entire genome. This technology has also been applied to colorectal cancers by others.[4)5)] In aCGH, which is used to detect an abnormal number of copies, it is very important to set a threshold. Thresholds have been determined uniformly based on the variation in the fluorescence ratio, but there have been situations in which this is not always sufficient. Therefore, the threshold-shift method was developed in order to compensate for this disadvantage. There are two types of methods for threshold shift: a method of exhaustively evaluating the gap between the determined values at a certain interval, and a method of evaluating the value itself as a threshold, which is analyzed for each clone in all cases, as was done in this study. When the aCGH chip and analysis software of Macrogen are used, the ratio of the fluorescence intensity is calculated to four decimal places as a value. When attempting to perform a similar calculation to four decimal places using the first method, an evaluation must be made in intervals of 0.0001. Therefore, when a gap between 0.5 and -0.5 is evaluated with an interval of 0.0001, it will be evaluated 10,000 x 4,030 times. As done in this study, when the value itself, which has been analyzed for each clone in all cases, is evaluated as a threshold, it will be evaluated as frequently as the number of cases x 4,030. The amount calculated for these 74 cases of colon cancer is 74 x 4,030 for the kai square test. This method was applied to an aCGH analysis of colon cancer in order to verify whether it is suitable.

Some of the clones that were detected using this method have been reported as chromosomal regions that are associated with colon cancer. An increased number of copies of clone 2748 (ZNF16, Tmp21-II, 8q24.3) was detected exclusively in Stage III/IV of colon cancer, but there are also many reports that this is related to biological characteristics such as transfers.[6)7)8)9)] The threshold in this clone was 0.2483, and a numeric value that is close to the threshold of 0.25 was usually used.

Clone 1102 (3q13.33) has been coded as STXBP5L, and this has been reported as likely to contain a cancer-associated gene.[10)] For example, in the experimental results using this aCGH, the cases that showed a loss of up to -0.15 included very few cases of Stages III & IV (1 in 13 cases), which supports the hypothesis that STXBP5L is a cancer-associated gene.

It has been reported that 7q11.23 for clone 5617, which was frequently detected as being abnormal in the Stage III/IV group, contains a cancer-associated gene.[11)] The loss of region 20p12.1, where clone 1509 exists, is commonly found in the cell lines of colon cancer.[12)] Considering that many cell lines are established from advanced cancer, it is hardly contradictory that most (16 in 17 cases) of the loss at 0.1182 or less is present in Stages III & IV.

It has been reported that 1q44 for clone 4684 is a fragile site of lymphocytes in patients with rectum cancer due to aphidicolin treatment.[13)]

Clone 2859, which exists in region 17p12, contains MAP2K4, which is reportedly associated with microsatellite instability and the prognosis of colon cancer.[14)15)16)] These reports do not contradict the result that most (18 in 21 cases) of the loss at up to -0.2667 in the experimental results was present in Stages III & IV.

As described above, these clones were selected using an arbitrarily set threshold, but there are some clones in regions that have also been reported in reference to colon cancer. In addition, there were no clones that provided a counterexample for $p<0.001$. This indicates the usefulness of this method when selecting a clone that has a difference between the two groups using the array CGH. It is expected that aCGH will be used clinically, and the method that has been developed herein should be useful.

Conclusions

With this method, the threshold of each BAC clone fluctuates automatically depending on the number of samples in order to enable the identification of a BAC clone correlated with the clinicopathological parameters. In other words, a convenient threshold is set so that the significant difference in a kai square test between two groups classified according to a certain clinicopathological parameter is maximized. Moreover, its usefulness has been verified based on 74 cases of colon cancer. When 74 cases of colon cancer were divided into 2 groups stage I/II and stage III/IV there was a significant difference ($P<0.001$) between the two groups only in the gain of the number of clones in region 8q24.3 when testing only with a threshold of ± 0.25. Using this method of analysis, the gain of the number of clones copied in region 7q11.23 and the loss of the number of clones copied in regions 20p12.1, 5q33.3, 17p12, and 9q33.1 were also observed in significant levels ($p<0.001$) in stage III/IV of colon cancer. Meanwhile, the gain of the number of clones copied in region 1q44 and the loss of the number of clones copied in regions 3q13.33, 14q12, 5p15.1, and 14q32.11 correlated with stage I/II colon cancer. The use of this method enabled the efficient identification of clones that were correlated with a specific parameter.

Authors' contributions

The authors declare no competing interests.

Acknowledgements

The study protocol was conducted under the approval of the Institutional Review Board for Human Use at the Yamaguchi University School of Medicine in 2004, and informed consent for this study was obtained from all patients.

References

1) Pinkel, D., Segraves, R., Sudar, D., Clark, S., Poole, I., Kowbel, D., Collins, C., Kuo, W.L., Chen, C., Zhai, Y., Dairkee, S.H., Ljung, B.M., Gray, J.W. and Albertson, D.G.: High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays. *Nat. Genet.*, **20** : 207-211, 1998.

2) Ng, G., Huang, J., Roberts, I. and Coleman, N.: Defining ploidy-specific thresholds in array comparative genomic hybridization to improve the sensitivity of detection of single copy alterations in cell lines. *J. Mol. Diagn.*, **8** : 449-458, 2006.

3) Frankenberger, C., Wu, X., Harmon, J., Church, D., Gangi, L.M., Munroe, D.J. and Urzúa, U.: WebaCGH: an interactive online tool for the analysis and display of array comparative genomic hybridisation data. *Appl. Bioinformatics*, **5** : 125-130, 2006.

4) Lassmann, S., Weis, R., Makowiec, F., Roth, J., Danciu, M., Hopt, U. and Werner, M.: Array CGH identifies distinct DNA copy number profiles of oncogenes and tumor suppressor genes in chromosomal- and microsatellite-unstable sporadic colorectal carcinomas. *J. Mol. Med.*, **85** : 293-304, 2007.

5) Nakao, K., Mehta, K.R., Fridlyand, J., Moore, D.H., Jain, A.N., Lafuente, A., Wiencke, J.W., Terdiman, J.P. and Waldman, F.M.: High-resolution analysis of DNA copy number alterations in colorectal cancer by array-based comparative genomic hybridization. *Carcinogenesis*, **25** : 1345-1357, 2004.

6) Van, Duin, M., Van, Marion, R., Vissers, K.J., Hop, W.C., Dinjens, W.N., Tilanus, H.W., Siersema, P.D. and Van, Dekken, H.: High-resolution array comparative genomic hybridization of chromosome 8q: evaluation of putative progression markers for gastroesophageal junction adenocarcinomas. *Cytogenet. Genome Res.*, **118** : 130-137, 2007.

7) Kim, J.H., Dhanasekaran, S.M., Mehra, R., Tomlins, S.A., Gu, W., Yu, J., Kumar-Sinha, C., Cao, X., Dash, A., Wang, L., Ghosh, D., Shedden, K., Montie, J.E., Rubin, M.A., Pienta, K.J., Shah, R.B. and Chinnaiyan, A.M.: Integrative analysis of genomic aberrations associated

with prostate cancer progression. *Cancer Res.*, **67** : 8229-8239, 2007.

8) Buffart, T.E., Coffa, J., Hermsen, M.A., Carvalho, B., Van, der, Sijp, J.R., Ylstra, B., Pals, G., Schouten, J.P. and Meijer, G.A.: DNA copy number changes at 8q11-24 in metastasized colorectal cancer. *Cell. Oncol.*, **27** : 57-65, 2005.

9) Saha, S., Bardelli, A., Buckhaults, P., Velculescu, V.E., Rago, C., St, Croix, B., Romans, K.E., Choti, M.A., Lengauer, C., Kinzler, K.W. and Vogelstein, B.: A phosphatase associated with metastasis of colorectal cancer. *Science*, **294** : 1343-1346, 2001.

10) Katoh, M. and Katoh, M.: Identification and characterization of human LLGL4 gene and mouse Llgl4 gene in silico. *Int. J. Oncol.*, **24** : 737-742, 2004.

11) Takekawa, M., Itoh, F., Hinoda, Y., Adachi, M., Ariyama, T., Inazawa, J., Imai, K. and Yachi, A.: Chromosomal localization of the protein tyrosine phosphatase G1 gene and characterization of the aberrant transcripts in human colon cancer cells. *FEBS Lett.*, **339** : 222-228, 1994.

12) Davison, E.J., Tarpey, P.S., Fiegler, H., Tomlinson, I.P. and Carter, N.P.: Deletion at chromosome band 20p12.1 in colorectal cancer revealed by high resolution array comparative genomic hybridization. *Genes Chromosomes Cancer*, **44** : 384-391, 2005.

13) Tunca, B., Egeli, U., Zorluoğlu, A., Yilmazlar, T., Yerci, O. and Kizil, A. : The expression of fragile sites in lymphocytes of patients with rectum cancer and their first-degree relatives. *Cancer Lett.*, **152** : 201-209, 2000.

14) Kim, W.H., Lee, H.W., Park, S.H., Kim, Y.I. and Chi, J.G.: Microsatellite instability in young patients with colorectal cancer. *Pathol. Int.*, **48** : 586-594, 1998.

15) Liu, X.P., Kawauchi, S., Oga, A., Sato, T., Ikemoto, K., Ikeda, E. and Sasaki, K.: Chromosomal aberrations detected by comparative genomic hybridization predict outcome in patients with colorectal carcinoma. *Oncol. Rep.*, **17** : 261-267, 2007.

16) Hermsen, M., Postma, C., Baak, J., Weiss, M., Rapallo, A., Sciutto, A., Roemen, G., Arends, J.W., Williams, R., Giaretti, W., De, Goeij, A. and Meijer, G.: Colorectal adenoma to carcinoma progression follows multiple pathways of chromosomal instability. *Gastroenterology*, **123** : 1109-1119, 2002.