Identification of chromosomal regions with DNA copy number aberrations associated with node metastasis of colorectal adenocarcinomas based on the array CGH profiles

Motonao Nakao[a], Shigeto Kawauchi[a],

Tomoko Furuya[a], Tetuji Uchiyama[b], Jun Adachi[b],

Takae Okada[a], Kenzo Ikemoto[a], Atsunori Oga[a], and Kohsuke Sasaki[a§]

[a]*Department of Pathology, Yamaguchi University Graduate School of Medicine, 1-1-1, Minami-Kogushi, Ube, Yamaguchi 755-8505, Japan*

[b]*Department of Surgery, Iwakuni Medical Center, Iwakuni 740-0021, Japan.*

[§]Correspondence:

Email addresses: kohsuke@yamaguchi-u.ac.jp

This is supported by grants from Japan Society for the Promotion of Science (19390102) and NEDO

Abstract

It is important to estimate biological characteristics of tumor, including the nodal status at the time of diagnosis for optimal treatment for individual cancer patients.

Array-based comparative genomic hybridization (aCGH) was performed for 77 sporadic colorectal adenocarcinomas using a chip spotted with 4,030 BAC clones. The nodal status was compared with an array CGH profiles depicted using a combination of decision-tree classifier and a Self-Organizing Map (SOM) analysis.

Node metastasis was not detected in any of the six poorly differentiated adenocarcinomas with a 3q loss. A SOM analysis following the decision-tree classification of the aCGH data allowed for the differentiation in chromosomal regions between high- and low-level decreases in the DNA copy number. Node metastasis was detected in all five tumors with the high-level decrease in DNA copy number at Xp, irrespective of the histological type. Node metastasis was also found exclusively in six tumors with an increase in DNA copy number at the chromosomal region between 11q13.3 and 11q22.3.

Chromosomal regions with copy number aberrations linked to nodal metastasis were identified more collectively by the combination of the decision-tree classifier and a SOM analysis than by the conventional analysis method in aCGH analysis.

Introduction

Colorectal cancer is one of the most frequent cancers in the world, and many people die of the disease. Cancer is the leading cause of death in developed countries. In principle, the prognosis of patients is much better in early cancers than in advanced cancers. Lymph node metastasis that is frequent in advanced cancers greatly affects the prognosis of cancer patients. It is important to determine the nodal status in advance of treatment for determining optimal therapeutic modality in individual patients. This is a practical issue, since minimally invasive procedures are recommended to achieve a high quality of life in cancer treatment, and overzealous treatment should be avoided. It is thought that the biological characteristics of a tumor primarily depend on the underlying genomic aberrations of the cancer cells and that the estimation of the biological characteristics may require comprehensive analysis of genomic alterations in cancer cells [1]. A comparative genomic hybridization (CGH) seems to meet the requirement. In particular, a high-resolution method, array-based CGH (aCGH), provides the high resolution analysis of DNA copy number aberrations in the whole genome of cancer cells, and aCGH reveals genomic changes associated with tumor development and progression [2-3]. In addition, the technology has been applied to surgically removed tissue specimens to identify genomic changes linked with clinicopathological features of specific types of cancer [4-6]. However, such an attempt is still limited so that the association between genomic changes and clinicopathological characteristics is still not clear for most of tumors. In this context, the clinical applicability of aCGH has not been defined. The present study identified the chromosomal regions with DNA copy number aberrations associated with nodal

status in colorectal cancer using a sophisticated method analysis of the aCGH profiles that is a combination of decision-tree classifier and a self-organizing map (SOM) analysis method.

Materials and Methods

1. *Sample preparation*

Seventy-seven surgically removed sporadic colorectal adenocarcinoma specimens were used for aCGH. The patients included 41 males and 36 females. The colorectal cancers were histologically classified into five well differentiated (tub1), 52 moderately differentiated (tub2), 14 poorly differentiated (por) and six mucinous carcinomas (muc). The histological classification of colorectal adenocarcinoma was made according to the General Rules for Clinical and Pathological Studies on Cancer of the Colon, Rectum and Anus (Japanese Society for Cancer of the Colon and Rectum). None of the patients received either irradiation or chemotherapy prior to surgery. In this series, 39 cases had lymph node metastasis.

High molecular weight genomic DNA was extracted from microdissected tumor tissue specimens using a DNA extraction kit. (Dneasy Tissue kit, QIAGEN SCIENCES, Valencia, CA). Sample DNA (500 ng, BioPrime DNA Labeling System, Invitrogen, Carlsbad, CA) was labeled with FluoroLink^{TM} Cy5-dCTP (Perkin Elmer, Wellesley, MA). Reference DNA (Human Genomic DNA, Promega, Madison, WI) is labeled with FluoroLink^{TM} Cy3-dCTP (Perkin Elmer). These DNAs were applied to the CGH array slides (MacArray Karyo4000, Macrogen Inc, Seoul Korea). The array was spotted with 4,030 human bacterial artificial chromosome (BAC) clones that covered the human whole genome at an average interval of 0.83Mbp. Images of 16-

bit fluorescence intensity for spots were captured from each array slide with a Gene Pix 4000A scanner (Axon Instruments, Burlingame, CA) and then the ratio value was calculated using a MAC Viewer (Macrogen Inc, Seoul Korea). Inadequate spots were flagged by manual inspection. All fluorescence intensity ratios were converted to log base 2. The ratio data were treated with Adaptive Weights Smoothing (AWS) by use of the GLAD program developed by the R package [7-8].

2. *Data analysis*

In order to identify chromosomal regions with different genomic aberrations between carcinomas with and without nodal metastasis, initially a classification of the aCGH profile was made by the decision-tree model J48 classifier attached to WEKA [9] (http://www.cs.waikato.ac.nz/ml/weka/). When the discrimination performance was incomplete in the first classification, classification procedure was repeated after eliminating the clones (spots) leading to a misclassification from the data to improve the analysis accuracy. Subsequently, an unsupervised classification, SOM [10], of the array CGH profiles in each chromosome was applied to chromosomal regions identified by the decision-tree model to elucidate the relationship of genomic aberrations to lymph node metastasis.

Results

The conventional aCGH analysis revealed a frequent gain of 20q, specially the gain of 20q11.21 was detected in 70% of colorectal cancers. A loss of 18q was also frequently observed; especially the loss of 18q23 was detected in 68% of cancers (**Fig.**

1). In addition, gains of 7p, 8q and chromosome 13, and losses of 8p and 17p were distinct.

In order to identify chromosomal regions roughly linked to nodal status, first the decision-tree model was applied to the aCGH data. This procedure allowed identification of chromosomal regions and their copy numbers linked to node metastasis of colorectal cancers. Chromosomal regions 11q13.3, Xp, 15q26.2 – q26.3, 2q37.1, Xq23, 8q24.3, and 8p23.3 together with their copy number were identified as makers of node metastasis in 77 colorectal cancers (**Fig. 2(A)**). However, five tumors were misclassified at chromosomal regions Xq23 and 8q24.3 by the first analysis. Reanalysis was performed after eliminating these regions from the data, and consequently ten chromosomal regions were selected as node metastasis markers, 11q13.3, 15q26.2-q26.3, Xp22.31, 1q23.1, 8p12, 7q11.21, 9p21.1-13.3, 13q34, 17q22, and 1p12-p11.2 (Fig. 2(B)). The relationship between DNA copy number aberrations and node metastasis was examined for each histological subtype. This series contained 52 moderately differentiated adenocarcinomas, and of these tumors 26 (50%) exhibited node metastasis. In moderately differentiated adenocarcinomas, node metastasis was detected in all of the 18 cancers with DNA copy number of $\leqq$0.1171 at 9p34.3, of >0.0939 at 7q36.3 and of >-0.38055 at 15q22.2 (Fig. 2 (C)). The copy number aberrations were detected in 18 (69.2%) of 26 cancers with node metastasis. In contrast, no node metastasis was detected in all of 16 cancers with DNA copy number of $\leqq$0.1171 at 9p34.3, of $\leqq$0.0939 at 7q36.3, of $\leqq$0.1067 at 11p12 and of $\leqq$ 0.06445 at 10q26.13. The repeated decision-tree analysis after eliminating clones leading to misclassification from the data did not significantly improve the diagnostic accuracy (**Fig. 2(D)**). In this series, there were 14 poorly differentiated adenocarcinomas of which seven (50%) showed node metastasis. In poorly

differentiated adenocarcinomas, node metastasis was detected in all of six cancers with a DNA copy number of >0.00315 at 3q21.1 (**Fig. 2(C)**). Namely, 85.7% (6/7) of poorly differentiated adenocarcinomas with node metastasis showed DNA copy number of >0.00315 at 3q21.1.

Subsequently, an SOM analysis was performed for the six chromosomal regions identified, 3q, 7q, 10q, 11q, 15q and Xp, by the decision-tree analysis. For poorly differentiated adenocarcinomas (por), a 2 x 2 grid was used (**Fig. 3 (A)**), and for moderately differentiated adenocarcinomas, a 3 x 3 was used (**Fig. 3 (C)**). For all 77 cases, a 5 x 5 grid was used (**Fig. 3 (B), (D), (E), and (F**)).In poorly differentiated adenocarcinoma, a profile representing 3q loss, as shown in **#4 of Figure 3 (A)**, was associated with lymph-node-negative. Of 14 poorly differentiated adenocarcinomas six (42.9%) showed this profiles, and all of them were devoid of node metastasis. In this histological subtype cancer, no node metastasis in seven cancers, and six (85.7%) of these cancers exhibited the profile. With respect to chromosomal region 7q in 77 cancers, however, no SOM profiles linked to node metastasis were identified (**Fig. 3(B)**). All of the six moderately differentiated adenocarcinomas with the whole 10q loss shown in **#6 of Figure 3 (C)** had node metastasis. Node metastasis was detected in all of six cancers with a copy number gain of 11q **(#5 in Fig. 3 (D)**). In addition, node metastasis was detected in seven of 10 cancers with the low level loss of 15q (#16, #17, #23, and #24 in **Fig. 3 (E)**). Node metastasis was detected in all of five cancers with the profile shown in **#1 of Figure 3 (F)**, while no node metastasis was detected in three cancers with the profile shown in #6 in which the copy number gain extended the entire length of Xp.

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 [11-12]. The 20q gain and 18q loss that were the most frequent gain and loss, respectively in this study were also frequently detected in other studies of colorectal cancers [11-15]. In addition, gains of 7p, 8q and chromosome 13, and losses of 8p and 17p were shown by other investigators who examined Japanese specimens as well as by the present study [15-17]. These observations indicate the reliability of the present a CGH data.

aCGH technology is currently advancing a transition from being purely a research tool to being used in clinical diagnostics. However, it has been not yet applied to cancer diagnosis [18]. This is partly due to cumbersome procedures of data analysis inherent to microarray technologies. In this study, an analytical method designed in this laboratory was used to identify genomic markers linked to the nodal status of colorectal cancer patients [12]. The method was a two-step procedure; first, the decision-tree model identified chromosomal regions that were linked to node metastasis. Thereafter, the SOM method was applied to the chromosomal regions 3q, 7q, 10q, 11q, 15q and Xp, which were identified by the decision-tree classifier. The SOM analysis combined with the decision-tree method allowed categorization of aCGH profiles for each chromosome with the objective of identifying the relationship between copy number changes and node metastasis.

Generally, colorectal cancers with metastasis to lymph nodes and/or other organs show more complex genomic aberrations than those without metastasis [19]. The copy number increase in 20q was previously reported to be associated with node metastasis of colorectal cancers [20]. However, the 20q gain may be practically

inadequate as a marker of node metastasis, because the 20q gain was highly frequent in colorectal cancers in this study as well as in others [12, 21-22]. Indeed, an association of copy number aberrations of 20q with nodal status was not evident in this study. DNA copy number aberrations in chromosomal regions 3q, 10q, 11q, 15q and Xp were linked to node metastasis. The association of copy number changes at these chromosomal regions with node metastasis has not been reported. The 3q loss represented node metastasis in poorly differentiated adenocarcinomas. In addition, the extent of copy number aberrations measured by aCGH should be emphasized on the estimation of the nodal status. The high-level decrease in copy number at Xp was associated with node metastasis, whereas the low-level change was not. The low-level decrease in DNA copy number at 10q or 15q was frequent in cancers with node metastasis. The mechanisms by which the copy number changes affect a phenotype of the tumor remain to be elucidated. The present aCGH data analysis procedure allowed the identification of copy number aberrations linked to the nodal status. In the studies like this, however, we must keep in mind that although this method allows high-accuracy identification of a group with nodal metastasis, the group may be a minor population of tumors with nodal metastasis.

The detection of copy number changes linked to node metastasis is useful not only for elucidating the genetic mechanisms of node metastasis but also for estimating node metastasis before treatment. The present study revealed multiple chromosomal regions where copy number aberrations were closely linked to node metastasis. It would therefore not be surprising to note that no single aberration exists which is common to all tumors with node metastasis. This also suggests that there are multiple pathways of node metastasis as well as of cancer development in the colon and rectum [23]. Node metastasis may largely depend on the tumor development pathway in each

cancer. It is also important to identify genomic alterations linked to patient prognosis, and thus patients are being followed closely.

Conclusions

The aCGH data analysis based on the combination of a decision-tree model and an SOM analysis thus made it possible to identify the aCGH profiles linked to the nodal status. Chromosomal regions 3q, 10q, 11q, 15q and Xp of which the DNA copy number aberrations were linked to node metastasis were identified.

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. Kallioniemi A, Kallioniemi OP, Sudar D, Rutovitz D, Gray JW, Waldman F, Pinkel D. Comparative genomic hybridization for molecular cytogenetic analysis of solid tumors. Science 1992; 258 : 818-821.
2. Pinkel D, Segraves R, Sudar D, Clark S, Poole I, Kowbel D, Collins C, Kuo WL, Chen C, Zhai Y, Dairkee SH, Ljung BM, Gray JW, Albertson DG. High

resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays. Nat Genet 1998; 20 : 207-211.

3. Pollack JR, Perou CM, Alizadeh AA, Eisen MB, Pergamenschikov A, Williams CF, Jeffrey SS, Botstein D, Brown PO. Genome-wide analysis of DNA copy-number changes using cDNA microarrays. Nat Genet, 1999; 23 : 41-46.
4. Mehta KR, Nakao K, Zuraek MB, Ruan DT, Bergsland EK, Venook AP, Moore DH, Tokuyasu TA, Jain AN, Warren RS, Terdiman JP, Waldman FM. Fractional genomic alteration detected by array-based comparative genomic hybridization independently predicts survival after hepatic resection for metastatic colorectal cancer. Clin Cancer Res 2005; 11:1791-7.
5. Knosel T, Schluns K, Dietel M, Petersen I. Chromosomal alterations in lung metastases of colorectal carcinomas: associations with tissue specific tumor dissemination. Clin Exp Metastasis 2005; 22:533-8.
6. Hirasaki S, Noguchi T, Mimori K, Onuki J, Morita K, Inoue H, Sugihara K, Mori M, Hirano T. BAC clones related to prognosis in patients with esophageal squamous carcinoma: an array comparative genomic hybridization study. Oncologist 2007; 12:406-17.
7. Polzehl J, Spokoiny VG. Adaptive Weights Smoothing with applications to image restoration. J Roy Stat Soc B 2000; 2 : 335-354.
8. Hupe P, Stransky N, Thiery JP, Radvanyi F, Barillot E. Analysis of array CGH data: from signal ratio to gain and loss of DNA regions. Bioinformatics 2004; 20 : 3413-3422.
9. Frank E, Hall M, Trigg L, Holmes G, Witten IH. Data mining in bioinformatics using Weka. Bioinformatics 2004; 20 : 2479-2481.

10. Dietzsch J, Gehlenborg N, Nieselt K. Mayday a microarray data analysis workbench. Bioinformatics 2006; 22 : 1010-1012.
11. Lassmann S, Weis R, Makowiec F, Roth J, Danciu M, Hopt U, 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 2007; 85: 293-304.
12. Nakao K, Mehta KR, Fridlyand J, Moore DH, Jain AN, Lafuente A, Wiencke JW, Terdiman JP, Waldman FM. High-resolution analysis of DNA copy number alterations in colorectal cancer by array-based comparative genomic hybridization. Carcinogenesis 2004; 25:1345-57.
13. Liu XP, Sato T, Oga A, Ikemoto K, Kawauchi S, Ikeda E, Sasaki K. Two subtypes of mucinous colorectal carcinoma characterized by laser scanning cytometry and comparative genomic hybridization. Int J Oncol 2004; 25: 615-21.
14. Ghadimi BM, Grade M, Liersch T, Langer C, Siemer A, Füzesi L, Becker H. Gain of chromosome 8q23-24 is a predictive marker for lymph node positivity in colorectal cancer. Clin Cancer Res 2003; 9: 1808-14.
15. Kurashina K, Yamashita Y, Ueno T, Koinuma K, Ohashi J, Horie H, Miyakura Y, Hamada T, Haruta H, Hatanaka H, Soda M, Choi YL, Takada S, Yasuda Y, Nagai H, Mano H. Chromosome copy number analysis in screening for prognosis-related genomic regions in colorectal carcinoma. Cancer Sci. 2008; 99:1835-1840.
16. Liu XP, Kawauchi S, Oga A, Sato T, Ikemoto K, Ikeda E, Sasaki K. Chromosomal aberrations detected by comparative genomic hybridization

17. Nakao K, Shibusawa M, Ishihara A, Yoshizawa H, Tsunoda A, Kusano M, Kurose A, Makita T, Sasaki K. Genetic changes in colorectal carcinoma tumors with liver metastases analyzed by comparative genomic hybridization and DNA ploidy. Cancer. 2001;91:721-726.
18. Bejjani BA, Shaffer LG. Application of array-based comparative genomic hybridization to clinical diagnostics. J Mol Diagn 2006 ; 8: 528-33.
19. Al-Mulla F, Behbehani AI, Bitar MS, Varadharaj G, Going JJ. Genetic profiling of stage I and II colorectal cancer may predict metastatic relapse. Mod Pathol 2006; 19:648-58.
20. Nakao K, Shibusawa M, Ishihara A, Yoshizawa H, Tsunoda A, Kusano M, Kurose A, Makita T, Sasaki K. Genetic changes in colorectal carcinoma tumors with liver metastases analyzed by comparative genomic hybridization and DNA ploidy. Cancer 2001; 91 : 721-726.
21. Korn WM, Yasutake T, Kuo WL, Warren RS, Collins C, Tomita M, Gray J, Waldman FM. Chromosome arm 20q gains and other genomic alterations in colorectal cancer metastatic to liver, as analyzed by comparative genomic hybridization and fluorescence in situ hybridization. Genes Chromosomes Cancer 1999; 25:82-90.
22. Xiao XY, Zhou XY, Yan G, Sun MH, Du X. Chromosomal alteration in Chinese sporadic colorectal carcinomas detected by comparative genomic hybridization. Diagn Mol Pathol 2007; 16:96-103.
23. Ogino S, Goel A. Molecular Classification and Correlates in Colorectal Cancer. J Mol Diagn 2008; 10: 13-27.

Figures

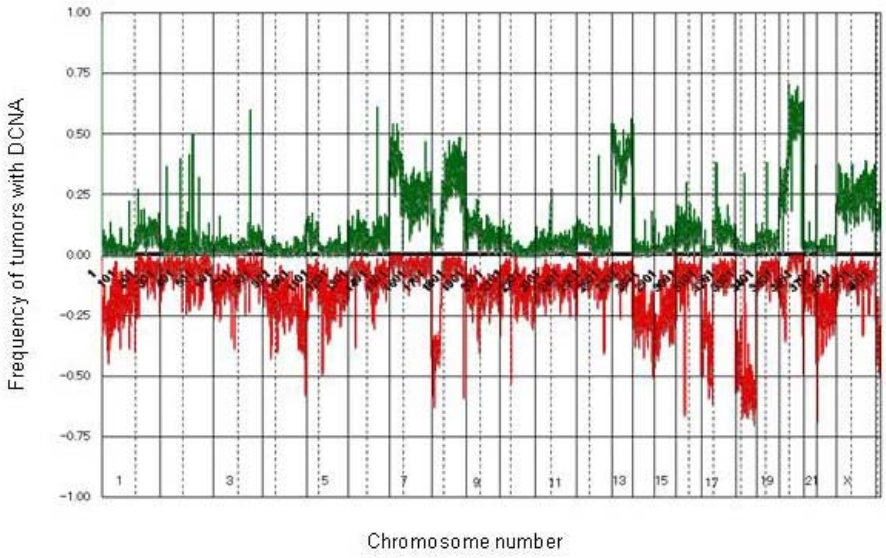

Fig. 1. Overall frequency of DNA copy number aberrations

Overall frequency of DNA copy number aberrations detected by aCGH for each BAC clone in 77 colorectal adenocarcinomas. The frequency of aberrations is depicted as a fraction of cases with DNA copy number gain or loss for 4000 BAC clones (the entire genome). The dots in the upper part of the profile indicate the frequency of tumors with DNA copy number gains, and dots in the lower part of the profile indicate the frequency of tumors with DNA copy number losses.

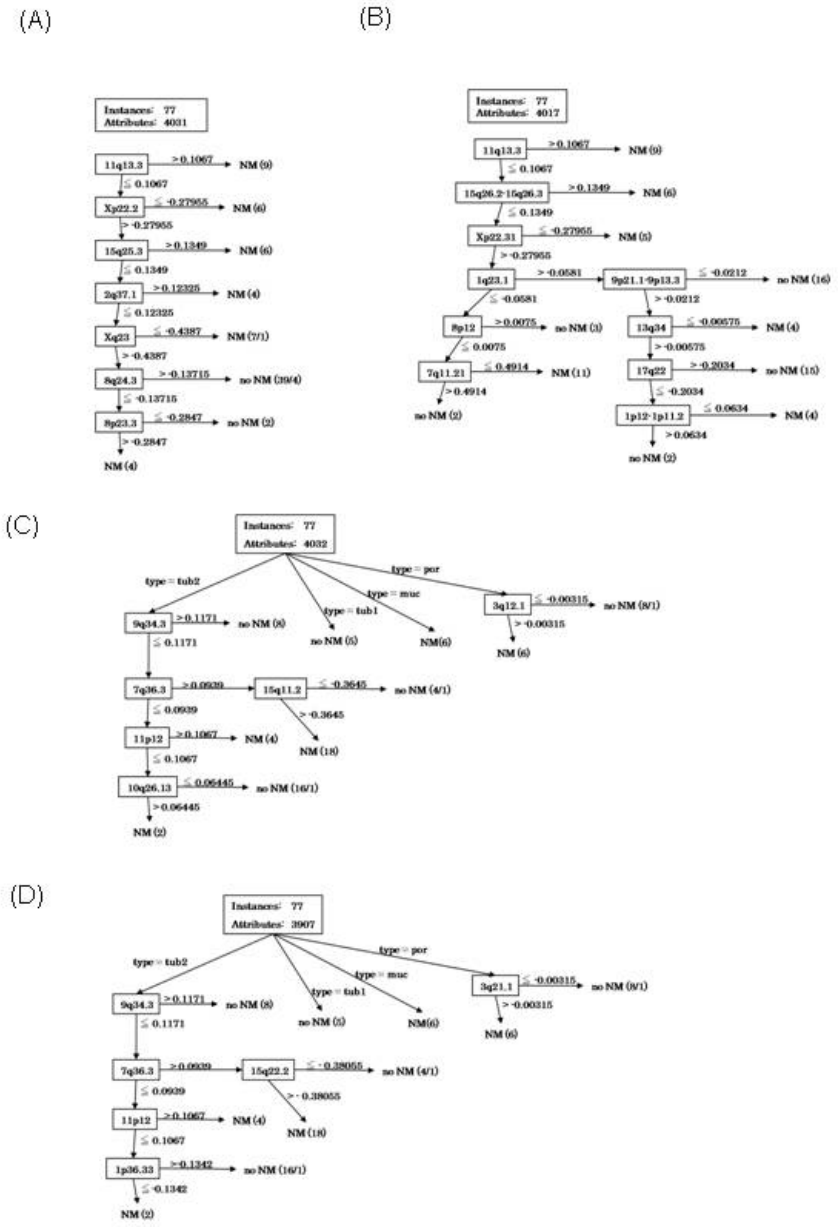

Fig. 2. A classification tree of the smoothed aCGH data

A classification tree of smoothed aCGH data was created, with a J48 classifier attached to WEKA. NM means the presence of nodal metastasis and the case number in parentheses is correctly classified, slash and false positive case number.

(A) The DNA copy number of a tumor at 11q13.3 is applied to the criterion of the first clone. If the copy number is larger than 0.1067 (log2), this tumor is estimated to have node metastasis. Nine tumors are classified into this group. When the copy

number of the tumor is not the case at the first clone, the second criterion should be checked. If the copy number of the tumor is equal or less than -0.27955 (log2) at Xp22.2, the tumor is considered node metastasis positive. Six tumors are classified into this group. When the copy number of the tumor at Xp22.2 does not meet the criterion of the second clone, the third clone is examined at 1p22.3. If the copy number of the tumor meets the third criterion, node metastasis is considered positive in the tumor. In this way, the copy number of the tumor is sequentially compared with that in each clone. The classification tree used all 4030 probes of aCGH. The classification performance was 93.5% (72/77).

(B) Reclassification is made after eliminating 14 clones leading to misclassification from data in the same way, the copy number of the tumor is sequentially compared with 11q13.3, 15q26.2-15q26.3, Xp22.31 and 1q23.1. If the copy number of tumor at 1q23.1 is equal or less than -0.0581 (log2), 8p12 is the next criterion. While the copy number of the tumor at 1q23.1 is larger than -0.0581(log2), 9p21.1-9p13.3 is the next criterion. The classification performance was 100% (77/77)

(C) The decision-tree classifier using 4030 clones applied to histological typing. At the first step, colorectal cancers are classified into four histological types (tub1, tub2, por and muc) based on histological differentiation. In this classifier, all tub1 are estimated to have no node metastasis, but all muc are estimated to have node metastasis. In tub2 tumors, five clones (9q34.3, 7q36.3, 11p12, 10q26.13, 15q11.2) are used for differentiation between tumors with and without node metastasis. If the histological differentiation type is por, one clone (3q21.1) can classify with or without node metastasis. The classification performance was 96.1% (74/77)

(D) The classification tree was used eliminating the 3905 probes and histological differentiation data without misreading the 125 probes. The classification performance was 96.1% (74/77).

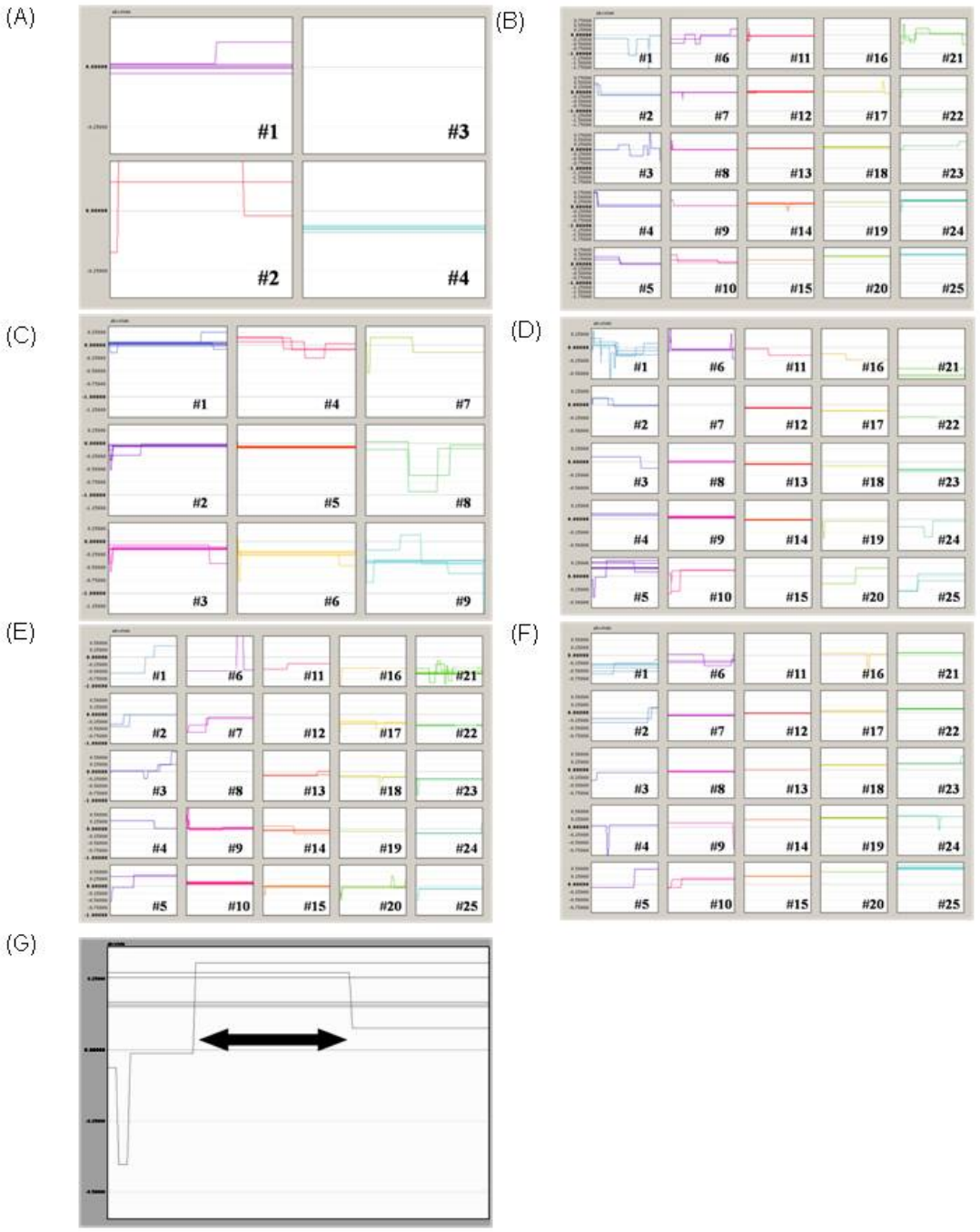

Fig. 3. Classification of the patterns of CGH profiles by SOM analysis

CGH profiles (A), (B), (C), (D), (E) and (F) corresponding to chromosomes 3q (2x2 grids, four blocks) in 14 poorly differentiated adenocarcinomas, 7q (5x5 grids, 25blocks) in 77 colorectal cancers, 10q (3x3 grids, nine blocks) in 77 colorectal cancers, 11q (5x5 grids, 25 blocks) in 77 colorectal cancers, 15q (5x5 grids, 25 blocks) in 77 colorectal cancers, and Xp (5x5 grids, 25 blocks) in 77 colorectal cancers, respectively. The location of SOM is shown as block numbers from #1 to #4 with 2x2 grids, from#1 to #9 with 3x3 grids and from #1 to #25 with 5x5 grids. In a block, the X-axis and Y-axis express each relative DNA copy number and chromosomal position. The position of clones for each chromosome in order from the telomer of the p arm to the telomere of the q arm Y-axis depcts fluorescence rations.

(G) The low-level gain of the DNA copy number at 11q is distinct in six tumors as shown in #5 of (E), and lymph node metastasis is detected in all of these six cases. The chromosome region with a low-level DNA copy number gain, as indicated by a double-headed arrow, is common for tumors with lymph node metastasis. Six tumors are categorized to have this pattern.