(7) and the projective adaptive resonance theory (PART) (8), prior to the application of mining algorithms.

In our previous study, we investigated the combinations of various filter and wrapper approaches and applied these combination methods to microarray data of acute leukemia and central nervous system tumors (CNS). Consequently, we showed that a combination method of the use of projective adaptive resonance theory and that of a boosted fuzzy classifier with the SWEEP operator method denoted PART-BFCS was the best among various combination methods for constructing on accurate model resulting in an accurate prediction. In this study, we applied this method to the analysis of expression profile data of esophageal cancer. In addition, the performances of BFCS or PART-BFCS with the U-test models, were investigated. The constructed PART-BFCS with the U-test or PART-BFCS models could accurately discriminate esophageal cancer patients with intramural metastases (IMs) from other esophageal cancer patients, BFCS with the U-test (U-test-BFCS) models could not.

It is necessary to select specific and essential marker genes for cancer classification and diagnosis. Minimum gene sets without false positive ones should be extracted. Therefore, various methods were compared under the condition of small inputs. We concluded that our method is the best under this condition for esophageal cancer analysis.

MATERIALS AND METHODS

Microarray analysis Gene expression profile data were obtained from 64 surgical specimens from esophageal cancer patients: 16 patients who had no lymph node metastases (O1), 6 patients who had lymph node metastases from one to four (O2), 29 patients who had over four lymph node metastases (O3), and 13 patients who had some IMs (see Table 1A). For RNA extraction, trained pathologists carefully excised bulk tissue samples from the main tumor, leaving a clear margin from the surrounding nontumorous tissue. Total RNAs extracted from the bulk tissue samples were biotin-labeled and hybridized to high-density oligonucleotide microarrays (Affymetrix Human Genome U95A Array) containing 12,600 probe sets representing 10,000 transcripts according to the manufacturer's instructions. The scanned data of the arrays were processed by Affymetrix Microarray Suite, which scaled the average intensity of all the genes on each array to a target signal of 1000.

Data processing As shown in Table 1B, the esophageal cancer data were partitioned into two data sets: 54 samples (42 non-IM and 12 IM) as a modeling data set for constructing the class prediction model (predictor) and 10 samples (9 non-IM and 1 IM) as a blind data set for evaluating the constructed predictor (10 blind data), and a leave-one-out cross-validation set (LOOCV data). We excluded genes expressed at a P call (meaning expression signal is present) of less than 10 in the 64 specimens. As a result, 8037 probes were selected in this preprocessing step. During the gene- filtering step, 1000 probes were selected using PART and the U-test, respectively, and then two types of BFCS, namely, BFCS-1 and BFCS-1,2 were used in the modeling step as wrapper approaches. For comparison, conventional modeling methods without filtering, namely, weighted voting (WV) (7) and k-nearest neighbor (kNN), were also used.

kNN method The k-nearest neighbor (kNN) method is based on a distance function for pairs of tumor samples, such as Euclidean distance. kNN proceeded as follows to classify blind data set observations on the basis of the modeling data set. For each patient in the blind data set (i) the k closest neighbors in the modeling data set were found, and (ii) class was predicted by majority vote; that is, the class that is most common among those k neighbors was chosen. The number of neighbors (k=3) was used because a similar cross-validation accuracy of models was obtained in the modeling data set for various ks.

WV method WV was originally proposed by Golub *et al.* (7) to manage microarray data. The weight of each gene was calculated using signal-to-noise statistic. The linear models of one gene were assembled by gene weight.

Model construction with parameter selection The parameter increasing method (PIM) (9) was used to select input combinations for the construction of kNN and WV models. This was performed as follows.

First, we predicted the class (IM or non-IM) of each sample using the prediction model with a single input. Prediction models for each probe were constructed in series, and all the probes were ordered on the basis of the accuracy of the constructed models. In the next step, the probe with the highest accuracy was used to construct a combination model.

Second, we selected a partner probe for the probe selected in the first step to increase prediction accuracy. To accomplish this, we

TABLE 1. List of esophageal cancer patients

A. All patients

Stage of metastasis	Description	Number of patients
O1	Lymph node metastases = 0	16
O2	4 ≥ Lymph node metastases ≥ 1	6
O3	Lymph node metastases > 4	29
IM	Intramural metastases (IM)	13
	Total	64

B. Divided data set

Data set name	Stage of metastasis	Content of data blocks		Number of data blocks
		Number in the modeling data	Number in the blind data	
Blind 10 data	Non-IM (O1, O2, O3)	42	9	1
	IM	12	1	
Leave-one-out Cross-validation (LOOCV) data	Non-IM	51	0	13
	IM	12	1	
	Non-IM	50	1	51
	IM	13	0	

constructed a 2-input model in which a ranked probe was designated input 1, and input 2 (partner probe) was selected to provide the highest training accuracy while applying kNN (or WV) and PIM to the analysis of the modeling data. By repeating this step, an optimum combination of $N_{attribute}$ candidate probes was identified for use as input probes in the model construction. $N_{attribute}$ was defined as ten in this study.

Finally, combinations of $N_{attribute}$ probes, *i.e.*, from the first to the $N_{attribute}$th probes were evaluated. We constructed $N_{attribute}$ predictor models, beginning with one input using only the first-selected probe to $N_{attribute}$ inputs using all the $N_{attribute}$ probes. The performance of the prediction models was evaluated by applying them to the analysis of the blind data set.

For the two data sets, the genes with the 1st to the 10th highest accuracies were used as the first inputs for the construction of the 10 combination models by PIM.

BFCS method Boosting was proposed by Schapire (10), and thus far, several derivative boosting algorithms (11–13) have been developed. Boosting is useful for class prediction using high-dimensional inputs and very fast algorithms.

In our previous study, we developed a boosted fuzzy classifier with the SWEEP operator method (BFCS) (5) on the basis of AdaBoost (11), which is the most basic boosting algorithm. This method enables the evaluation of the reliability of the predictions for each patient. However, it is difficult to evaluate the reliability of the predicted results of conventional boosting.

A BFCS model is composed of type I fuzzy neural network (FNN) models (14). In this study, 1- or 2-input FNN models were used as weak learners in the BFCS model, and they were combined with connection weights, which were determined using the AdaBoost algorithm. BFCS has two types, BFCS-1 and BFCS-1,2. A BFCS-1 model is composed of 1-input FNN models (5). On the other hand, BFCS-1,2 is composed of 1- or 2-input FNN models (5). BFCS-1,2 can used for analyzing the interaction between two inputs, because this method can includes 2-input FNN models.

PART-BFCS Previously, we developed and combined the use of the projective adaptive resonance theory (PART) as a gene filtering method and that of a boosted fuzzy classifier with the SWEEP operator method (BFCS) as a modeling method. In the resulting method PART-BFCS, PART first preselects the genes that show small variances within a class. Then, BFCS rapidly selects these genes to build a highly accurate and reliable predictor.

PART has two important parameters, vigilance and distance. Vigilance was optimized so that modeling samples clustered well. Distance was used to control the number of extracted genes. The genes extracted by PART showed a low standard deviation (SD) in the low-gene–expression-level class. The predictor using genes with a low SD in low class showed a high performance (8).

In BFCS, 1- or 2-input FNN models based on the neural network and fuzzy logic were used as weak learners. The BFCS models constructed using only 1-input FNN models were defined as a BFCS-1 model, and those constructed using 1- or 2-input FNN models were defined as a BFCS-1,2 model in our previous study.

RESULTS AND DISCUSSION

Selection of BFCS type and complexity of esophageal cancer data for the classification of IM and non-IM

BFCS-1 is effective for analyzing many gene expression profiles, such as those of acute leukemia, central nervous system tumors (CNS), and soft tissue sarcomas (unpublished data). BFCS-1 without screening was applied to the analysis of the modeling data of esophageal cancer shown in Fig. 1. Figure 1 shows training curves against the number

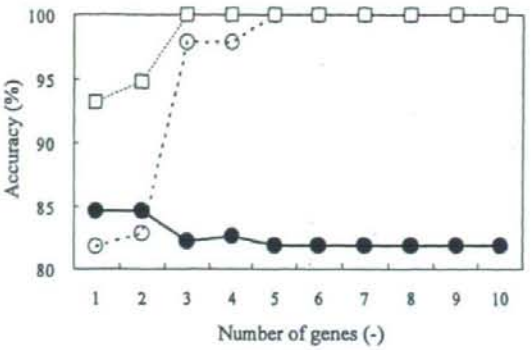

FIG. 1. Training curves of BFCS-1 without screening for modeling data of 10 blind data. The training curves were developed using average training accuracy from 10 combination models constructed by BFCS-1. The solid line with filled circles is the training curve for the esophageal cancer data. The dashed line with open circles is the curve for the acute leukemia data. The dashed line with open squares is the curve for the central nervous system (CNS) tumor data. The leukemia and CNS data were obtained from the website http://www.broad.mit.edu/cgi-bin/cancer/datasets.cgi.

of genes. The solid line indicates the training curve for the esophageal cancer data. The dashed lines indicate the modeling results for other cancer data, namely, the acute leukemia, and CNS data. The training curve result obtained by the BFCS-1 expressed underfitting of the esophageal cancer data, and a training curve result of 100% was achieved for the data of the other two cancers. This result implies that the esophageal cancer data were very complex. Therefore, BFCS-1,2 was used in this study, because it is more effective than BFCS-1 in the cases in which the relationships of the attributes provided and its output are highly complex.

Comparison of performances of BFCS with filtering methods with those of other methods The performances of BFCSs with filtering methods as models were investigated, namely, BFCS with PART (PART-BFCS), BFCS with the U-test (U-test-BFCS), and BFCS with PART and the U-test (PART-BFCS with U-test). For comparison, the predictors of two conventional methods, namely, WV and kNN, were constructed. The performances of the predictors were compared in terms of accuracy using a blind data set that was not used for modeling. By using 10 combination models, the average accuracy for the blind data set was calculated for the two data sets, namely, 10 blind and LOOCV data.

Results of LOOCV data are shown in Table 2. The results show that the average accuracy of 6-input PART-BFCS with the U-test models is the highest. The average accuracies of the BFCSs with filtering methods were higher than those of two conventional methods, namely, WV and kNN. However, U-test-BFCS models showed a very low sensitivity.

Results of 10 blind data are shown in Table 3. The results show that the average accuracy of 10-input PART-BFCS with the U-test methods is the highest and that the average accuracies of models for BFCS with filtering methods were higher than those of the conventional methods. However, U-test-BFCS model also shows a very low sensitivity.

A comparison of PART-BFCS and PART-BFCS with the

TABLE 2. Comparison of performances of various methods for LOOCV data

	Method (–)	Inputs (–)									
		1	2	3	4	5	6	7	8	9	10
Accuracy (%)	BFCS with PART and U-test	–	75.0	–	75.8	–	80.9[a]	–	78.8	–	80.3
	BFCS with PART	–	76.4	–	75.8	–	77.2	–	77.3	–	78.1
	BFCS with U-test	–	65.5	–	68.6	–	73.0	–	73.0	–	76.1
	kNN	74.7	70.0	70.8	70.2	71.3	69.5	70.3	68.1	69.4	69.1
	WV	61.3	64.1	66.1	69.8	63.0	62.2	63.6	65.9	65.9	64.7
Sensitivity (%)	BFCS with PART and U-test	–	15.4	–	21.5	–	21.5	–	13.1	–	11.5
	BFCS with PART	–	16.2	–	25.4	–	16.9	–	13.1	–	6.2
	BFCS with U-test	–	2.3	–	3.8	–	0.0	–	0.0	–	0.0
	kNN	23.8	24.6	25.4	20.8	21.5	18.5	16.2	14.6	19.2	16.9
	WV	14.6	12.3	13.8	16.9	15.4	19.2	17.7	16.2	15.4	16.2
Specificity (%)	BFCS with PART and U-test	–	90.2	–	89.6	–	96.1	–	95.5	–	97.8
	BFCS with PART	–	91.8	–	88.6	–	92.5	–	93.7	–	96.5
	BFCS with U-test	–	81.6	–	85.1	–	91.6	–	91.6	–	95.5
	kNN	87.6	81.6	82.4	82.7	83.9	82.5	84.1	81.8	82.2	82.4
	WV	73.1	77.3	79.4	83.3	75.1	73.1	75.3	78.6	78.8	77.1

[a] The highest accuracy. – indicates that no models were constructed, because BFCS-1,2 method selected a 2-input weak learner consisting of two genes. Accuracy is the ratio of correctly predicted patients to total patients. Sensitivity is accuracy for IM patients. Specificity is accuracy for non-IM patients.

TABLE 3. Comparison of performances of various methods for 10 blind data

	Method (–)	Inputs (–)									
		1	2	3	4	5	6	7	8	9	10
Accuracy (%)	BFCS with PART and U-test	–	80.0	–	84.0	–	85.0	–	89.0	–	96.0[a]
	BFCS with PART	–	83.0	–	81.0	–	82.0	–	83.0	–	88.0
	BFCS with U-test	–	82.0	–	79.0	–	84.0	–	83.0	–	88.0
	kNN	72.0	74.0	72.0	80.0	77.0	75.0	78.0	76.0	73.0	69.0
	WV	66.0	67.0	57.0	60.0	65.0	62.0	70.0	65.0	61.0	64.0
Sensitivity (%)	BFCS with PART and U-test	–	50.0	–	60.0	–	80.0	–	80.0	–	80.0
	BFCS with PART	–	70.0	–	80.0	–	90.0	–	90.0	–	90.0
	BFCS with U-test	–	30.0	–	10.0	–	10.0	–	0.0	–	0.0
	kNN	20.0	0.0	10.0	10.0	10.0	10.0	20.0	20.0	20.0	0.0
	WV	30.0	40.0	20.0	10.0	50.0	30.0	30.0	0.0	20.0	40.0
Specificity (%)	BFCS with PART and U-test	–	83.3	–	86.7	–	85.6	–	90.0	–	97.8
	BFCS with PART	–	84.4	–	81.1	–	81.1	–	82.2	–	87.8
	BFCS with U-test	–	87.8	–	86.7	–	92.2	–	92.2	–	97.8
	kNN	77.8	82.2	78.9	87.8	84.4	82.2	84.4	82.2	78.9	76.7
	WV	70.0	70.0	61.1	65.6	66.7	65.6	74.4	72.2	65.6	66.7

[a] The highest accuracy. – indicates that no models were constructed, because BFCS-1,2 method selected a 2-input weak learner consisting of two genes. Accuracy is the ratio of correctly predicted patients to total patients. Sensitivity is accuracy for IM patients. Specificity is accuracy for non-IM patients.

U-test was performed using the accuracies of 100 models (2 data sets × 10 combination models × 5 types of input from 2 to 10). The P value was 0.022 and was calculated using the paired t-test. PART-BFCS with the U-test was superior to PART-BFCS for esophageal cancer data. These results indicate that PART is necessary for BFCS, because PART eliminates genes which hinder the prediction of BFCS. In addition, PART-BFCS with the U-test was the best method for analyzing esophageal cancer data.

Comparison of selected genes by PART-BFCS and PART-BFCS with U-test The average accuracy of 6-input PART-BFCS with the U-test models was the highest, as shown in Table 2. The detailed results of ten combination 6-input PART-BFCS with the U-test models were analyzed (data not shown). Results of the PART-BFCS were also analyzed, because this method had the second highest accuracy of the 6-input models. The results showed that the accuracies of all the models used are almost the same. However, sensitivity markedly differed between the models; the sensitivities ranged from 0.0% to 46.2% for PART-BFCS with the U-test models, and from 7.7% to 38.5% for PART-BFCS models. The variance in sensitivity was large, because the number of IM patients was very small in this study. Therefore, the highest sensitivity models among ten combinations for each method were selected for the following analysis; the no. 4 model for PART-BFCS with the U-test and the no. 5 model for PART-BFCS.

Actually, 99 and 121 independent genes (probe sets) were selected and the top 10 genes that were selected most frequently are shown in Table 4A. Table 4A shows that the gene *CDK6* was selected most and the gene *SIM2* was selected 2nd most for both models. *CDK6* is a well-known cell cycle regulation gene and is an important marker for cancer diagnosis (15–17). For 10 blind data, *CDK6* was also selected frequently, as shown in Table 5.

Next, we investigated the genes selected together with

TABLE 4. List of genes selected by 6-input BFCS with screening for LOOCV data

A. The selected genes

Model	Gene name	Genbank	Description	Number of times selected
No. 4 model of BFCS with PART and U-test	CDK6	X66365	Cyclin-dependent kinase 6	45
	SIM2	U80456	Single-minded homolog 2 (*Drosophila*)	27
	MYL6	M22919	Myosin, light polypeptide 6, alkali, smooth muscle and non-muscle	19
	TRIP6	AJ001902	Thyroid hormone receptor interactor 6	19
	C19orf2	AB006572	Chromosome 19 open reading frame 2	17
	FBXO21	AB020682	F-box only protein 21	13
	KCNJ15	Y10745	Potassium inwardly-rectifying channel, subfamily J, member 15	12
	ZNF3	X07290	Zinc finger protein 3 (A8-51)	11
	POLS	AB005754	Polymerase (DNA directed) sigma	11
	NFIB	AI222594	Nuclear factor I/B	10
No. 5 model of BFCS with PART	CDK6	X66365	Cyclin-dependent kinase 6	37
	SIM2	U80456	Single-minded homolog 2 (*Drosophila*)	28
	C19orf2	AB006572	Chromosome 19 open reading frame 2	18
	TRIP6	AJ001902	Thyroid hormone receptor interactor 6	16
	POLS	AB005754	Polymerase (DNA directed) sigma	13
	ERCC1	M13194	Excision repair cross-complementing rodent repair deficiency, complementation group 1 (includes overlapping antisense sequence)	13
	FZD5	U43318	Frizzled homolog 5 (*Drosophila*)	12
	ZNF3	X07290	Zinc finger protein 3 (A8-51)	12
	NFIB	AI222594	Nuclear factor I/B	10
	TIAL1	D64015	TIA1 cytotoxic granule-associated RNA binding protein-like 1	9

B. Genes selected together with *CDK6*

Model	Gene name	Genbank	Description	Number of times selected
No. 4 model of BFCS with PART and U-test	C19orf2	AB006572	Chromosome 19 open reading frame 2	17
	MYL6	M22919	Myosin, light polypeptide 6, alkali, smooth muscle and non-muscle	9
	FZD5	U43318	Frizzled homolog 5 (*Drosophila*)	4
	FBXO21	AB020682	F-box only protein 21	3
	GPA33	U79725	Glycoprotein A33 (transmembrane)	3
	TRIP13	U96131	Thyroid hormone receptor interactor 13	2
	TCF4	M74719	Transcription factor 4	2
No. 5 model of BFCS with PART	C19orf2	AB006572	Chromosome 19 open reading frame 2	18
	FZD5	U43318	Frizzled homolog 5 (*Drosophila*)	12
	TRIP13	U96131	Thyroid hormone receptor interactor 13	2

(A) The list of these genes was sorted by the number of times selected in the LOOCV (64-fold), and the top 10 genes are shown. Independent 99 and 121 genes (probe sets) were selected for each model, respectively. Except for the names of genes described, those of other 89 genes (probe sets) involved in no. 4 model and 111 genes (probe sets) involved in no. 5 model were omitted. (B) BFCS-1,2 consisted of 2-input FNN models concluding two genes. Only the genes selected two or more times are shown. Except for the names of genes described, those of other 5 genes (probe sets) involved in each no. 4 and no. 5 model were omitted.

CDK6, as shown in Tables 4B and 5. For 10 blind data, Table 5 showed that *FZD5* and *GPA33* were frequently selected together with *CDK6* gene. For LOOCV data, Table 4B showed that *C19orf2* and *FZD5* were also selected frequently.

Comparison of accuracy of 2-input models including those for *CDK6* with those of other models The performances of 1- or 2-input BFCS models were calculated and are shown in Table 6, such as those for *CDK6+C19orf2*, *CDK6+FZD5*, *CDK6+GPA33*, *CDK6*, *C19orf2*, *FZD5*, *GPA33*, *CDK6+SIM2*, and the negative control. The negative control indicates the average performance of 2-input models selected randomly 20,000 times. Table 6 shows that the accuracies and sensitivities of 2-input models, such as those for *CDK6+C19orf2*, *CDK6+FZD5*, and *CDK6+GPA33*, are very high. On the other hand, the sensitivities of 1-input models, such as those for *CDK6*, *C19orf2*, *FZD5*, and *GPA33*, were zero percent. The irrelevant 2-input models, namely, those for *CDK6+SIM2* and the negative control, showed low sensitivities. These results show that all the patients are classified as non-IM patients by all the 1-input models used, because the 1-input models could not be constructed correctly owing to the high complexity of these data. These results show that 2-input combinations of *CDK6*, such as *CDK6+C19orf2*, *CDK6+FZD5*, and *CDK6+GPA33* are very important.

IF-THEN rules extracted from BFCS model After modeling, the IF-THEN rules for esophageal cancer with

TABLE 5. List of genes selected by BFCS with screening methods for 10 blind data

Method	Inputs (–)	Order of selection	Combination no. 1	2	3	4	5	6	7	8	9	10
BFCS with PART and U-test	2	1	POLS	HMGN1	SPTAN1	FBXO21	SHARP	PC4	RSU1	RSU1	SIM2	HMGN1
			BIG1	PC4	MEST	SIM2	SIM2	SIM2	G2AN	SIM2	ATP6AP2	PCSK1
	4	2	DNASE1L1	DNASE1L1	FBXO21	DNASE1L1	DNASE1L1	DNASE1L1	STARD3	DNASE1L1	DNASE1L1	RSU1
			Unknown	Unknown	TRIP6	Unknown	Unknown	Unknown	RAGE	Unknown	Unknown	G2AN
	6	3	HMGN1	SEC24A	HMGN1	HMGN1	HMGN1	SEC24A	HMGN1	HMGN1	HMGN1	DNASE1L1
			PC4	BIG1	PC4	PC4	PC4	BIG1	PC4	PC4	PC4	SLC10A3
	8	4	FBXO21	CDK6[a]	CDK6[a]	CDK6[a]	CDK6[a]	ERCC1	DNASE1L1	CDK6[a]	CDK6[a]	SEC24A
			TRIP6	C19orf2	LRP5	GPA33[b]	GPA33[b]	OXCT	Unknown	GPA33[b]	GPA33[b]	BIG1
	10	5	SHARP	FBXO21	OAS1	SEC24A	SEC24A	CDK6[a]	SEC24A	SEC24A	SEC24A	Unknown
			SIM2	SIM2	NFIB	BIG1	BIG1	GPA33	BIG1	BIG1	BIG1	BTAF1
BFCS with PART	2	1	POLS	HMGN1	SPTAN1	C21orf25	FBXO21	DKFZp547K	ARCN1	ZNF294	SHARP	NMU
			BIG1	PC4	MEST	SIM2	SIM2	SIM2	SIM2	SIM2	SIM2	SIM2
	4	2	SAA1	CDK6[a]	FBXO21	DNASE1L1	DNASE1L1	DNASE1L1	DNASE1L1	DNASE1L1	DNASE1L1	DNASE1L1
			SIM2	MADH4	TRIP6	Unknown	Unknown	Unknown	Unknown	Unknown	Unknown	Unknown
	6	3	CDK6[a]	CCBP2	HMGN1	HMGN1	HMGN1	HMGN1	HMGN1	HMGN1	HMGN1	HMGN1
			FLJ31564	POLS	PC4	PC4	PC4	PC4	PC4	PC4	PC4	PC4
	8	4	HMGN1	Unknown	CDK6[a]	CDK6[a]	CDK6[a]	CDK6[a]	CDK6[a]	CDK6[a]	CDK6[a]	CDK6[a]
			PC4	PRSS3	FLJ31564	FZD5[b]	FZD5[b]	FZD5[b]	FZD5[b]	FZD5[b]	FZD5[b]	FZD5[b]
	10	5	FBXO21	TERF1	OAS1	SAA1	SAA1	SAA1	SAA1	SAA1	SAA1	SAA1
			MINA53	MMP9	NFIB	BIG1	BIG1	BIG1	BIG1	BIG1	BIG1	BIG1

[a] *CDK6*.
[b] Genes were selected together with *CDK6*.

TABLE 6. Comparison of prediction accuracies of genes frequently selected by BFCS

Used genes (–)	Number of input	Accuracy (%)	Sensitivity (%)	Specificity (%)
CDK6+C19orf2[a]	2	89.1	53.8	98.0
CDK6+FZD5[a,b]	2	84.4	76.9	86.3
CDK6+GPA33[b]	2	82.8	76.9	84.3
CDK6	1	79.7	0.0	100.0
C19orf2	1	79.7	0.0	100.0
FZD5	1	79.7	0.0	100.0
GPA33	1	79.7	0.0	100.0
CDK6+SIM2[c]	2	79.7	30.8	92.2
Nagative control[d]	2	78.8±1.4	0.6±2.9	98.7±1.8

Accuracies were calculated by BFCS for LOOCV data.
[a] Gene that was frequently selected with *CDK6* for LOOCV data.
[b] Gene that was frequently selected with *CDK6* for 10 blind data.
[c] Gene that was the frequently selected 2nd for LOOCV data.
[d] Two genes were randomly extracted from the genes never selected by PART-BFCS or PART-BFCS with the U-test methods, and the model was constructed by BFCS. This procedure was repeated for 20,000 times.

IM and non-IM were obtained from the models including *CDK6*. The IF-THEN rules were obtained as a matrices that are classified by the expression level of selected genes for three 2-input models (Fig. 2). Using these matrices, simple and excellent rules were obtained as follows. The first rule is that patients with low expression levels of *CDK6* and *C19orf2* are likely to be IM patients, as shown in Fig. 2A. Seven patients showed low expression levels of *CDK6* and *C19orf2* and all of them were IM patients, corresponding to 54% (7/13) of all the IM patients. The next rule is that patients with low expression levels of *CDK6* and *FZD5* are likely to be IM patients, as shown in Fig. 2B. Sixteen patients showed low expression levels of *CDK6* and *FZD5* and 10 of them were IM patients, corresponding to 77% (10/13) of all the IM patients. The third rule is that patients

A

IM/NonIM		*CDK6* Low	*CDK6* High
C19orf2	Low	7/0	0/13
	High	3/16	3/22

B

IM/NonIM		*CDK6* Low	*CDK6* High
FZD5	Low	10/6	2/30
	High	0/8	1/7

C

IM/NonIM		*CDK6* Low	*CDK6* High
GPA33	Low	10/7	3/29
	High	0/7	0/8

FIG. 2. IF-THEN rules including those for *CDK6*. Because each gene can be divided into either a high or a low group using fuzzy logic, this model comprised 4 (=2^2) fuzzy rules. Values on the left in each matrix indicate the number of IM patients. Values on the right indicate the number of non-IM patients. (A) For *CDK6* and *C19orf2*. (B) For *CDK6* and *FZD5*. (C) *CDK6* and *GPA33*.

with low expression levels of *CDK6* and *GPA33* are likely to be IM patients, as shown in Fig. 2C. Seventeen patients showed low expression levels of *CDK6* and *GPA33* and 10 of them were IM patients, corresponding to 77% (10/13) of all the IM patients. Non-IM or IM patients clustered at spe-

cific parts of the matrices.

In this study, we applied PART-BFCS, and PART-BFCS with the U-test to discriminate esophageal cancer patients with IM from those with non-IM. It was necessary that a specific type of BFCS, BFCS-1,2, was used, because the esophageal cancer data used were highly complex. PART-BFCS and PART-BFCS with the U-test models showed higher performances than WV and kNN. PART-BFCS with the U-test was superior to PART-BFCS. The genes including *CDK6* were found using our methods. Accurate IF-THEN rules were extracted. The genes selected in this study have a high potential as new diagnosis markers for esophageal cancer. These results indicate that these methods are new methods of marker gene selection for the diagnosis of cancer patients.

ACKNOWLEDGMENTS

We would like to thank the Hori Information Science Promotion Foundation for financial support.

This work was supported in part by the program for the promotion of Fundamental Studies in Health Sciences of the National Institute of Biomedical Innovation (NiBio) and a Grant-in-Aid for the Third Comprehensive 10-Year Strategy for Cancer Control (10 years from FY2004) and for Cancer Research (16–15) from the Ministry of Health, Labour and Welfare of Japan.

REFERENCES

1. **Parkin, D. M., Bray, F., Ferlay, J., and Pisani, P.:** Estimating the world cancer burden: Globocan 2000. Int. J. Cancer, **94**, 153–156 (2001).
2. **Igaki, H., Kato, H., Tachimori, Y., Sato, H., Daiko, H., and Nakanishi, Y.:** Prognostic evaluation for squamous cell carcinomas of the lower thoracic esophagus treated with three-field lymph node dissection. Eur. J. Cardiothorac. Surg., **19**, 887–893 (2001).
3. **Ando, T., Suguro, M., Hanai, T., Kobayashi, T., Honda, H., and Seto, M.:** Fuzzy neural network applied to gene expression profiling for predicting the prognosis of diffuse large B-cell lymphoma. Jpn. J. Cancer Res., **93**, 1207–1212 (2002).
4. **Guyon, I., Weston, J., Barnhill, S., and Vapnik, V.:** Gene selection for cancer classification using support vector machines. Mach. Learn., **46**, 389–422 (2002).
5. **Takahashi, H. and Honda, H.:** A new reliable cancer diagnosis method using boosted fuzzy classifier with a SWEEP operator method. J. Chem. Eng. Jpn., **38**, 763–773 (2005).
6. **Takahashi, H. and Honda, H.:** Prediction of peptide binding to major histocompatibility complex class II molecules through use of boosted fuzzy classifier with SWEEP operator method. J. Biosci. Bioeng., **101**, 137–141 (2006).
7. **Golub, T. R., Slonim, D. K., Tamayo, P., Huard, C., Gaasenbeek, M., Mesirov, J. P., Coller, H., Loh, M. L., Downing, J. R., Caligiuri, M. A., Bloomfield, C. D., and Lander, E. S.:** Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. Science, **286**, 531–537 (1999).
8. **Takahashi, H., Kobayashi, T., and Honda, H.:** Construction of robust prognostic predictors by using projective adaptive resonance theory as a gene filtering method. Bioinformatics, **21**, 179–186 (2005).
9. **Noguchi, H., Hanai, T., Honda, H., Harrison, L. C., and Kobayashi, T.:** Fuzzy neural network-based prediction of the motif for MHC class II binding peptides. J. Biosci. Bioeng., **92**, 227–231 (2001).
10. **Schapire, R. E.:** The strength of weak learnability. Mach. Learn., **5**, 197–227 (1990).
11. **Freund, Y. and Schapire, R. E.:** A decision-theoretic generalization of online learning and an application to boosting. J. Comput. Syst. Sci., **55**, 119–139 (1997).
12. **Friedman, J., Hastie, T., and Tibshirani, R.:** Additive logistic regression: a statistical view of boosting. Ann. Stat., **28**, 337–407 (2000).
13. **Freund, Y.:** An adaptive version of the boost by majority algorithm. Mach. Learn., **43**, 293–318 (2000).
14. **Horikawa, S., Furuhashi, T., and Uchikawa, Y.:** On fuzzy modeling using fuzzy neural networks with the back-propagation algorithm. IEEE Trans. Neural Netw., **3**, 801–806 (1992).
15. **Mendrzyk, F., Radlwimmer, B., Joos, S., Kokocinski, F., Benner, A., Stange, D. E., Neben, K., Fiegler, H., Carter, N. P., Reifenberger, G., Korshunov, A., and Lichter, P.:** Genomic and protein expression profiling identifies CDK6 as novel independent prognostic marker in medulloblastoma. J. Clin. Oncol., **23**, 8853–8862 (2005).
16. **Garcia, J. F., Camacho, F. I., Morente, M., Fraga, M., Montalban, C., Alvaro, T., Bellas, C., Castano, A., Diez, A., Flores, T., Martin, C., Martinez, M. A., Mazorra, F., Menarguez, J., Mestre, M. J., Mollejo, M., Saez, A. I., Sanchez, L., and Piris, M. A.:** Hodgkin and Reed-Sternberg cells harbor alterations in the major tumor suppressor pathways and cell-cycle checkpoints: analyses using tissue microarrays. Blood, **101**, 681–689 (2003).
17. **Henshall, S. M., Quinn, D. I., Lee, C. S., Head, D. R., Golovsky, D., Brenner, P. C., Delprado, W., Stricker, P. D., Grygiel, J. J., and Sutherland, R. L.:** Overexpression of the cell cycle inhibitor p16INK4A in high-grade prostatic intraepithelial neoplasia predicts early relapse in prostate cancer patients. Clin. Cancer Res., **7**, 544–550 (2001).

Array-based comparative genomic hybridization of circulating esophageal tumor cells

DAL HO KIM[1,4], MANABU MUTO[5], YOSHITAKA KUWAHARA[1], YUKIHIRO NAKANISHI[2], HIROSHI WATANABE[3], KAZUHIKO AOYAGI[1], KENJI OGAWA[5], TERUHIKO YOSHIDA[1] and HIROKI SASAKI[1]

Divisions of [1]Genetics and [2]Pathology, National Cancer Center Research Institute; [3]Department of Surgery, National Cancer Center Hospital, 1-1, Tsukiji 5-chome, Chuo-ku, Tokyo 104-0045; [4]Medical Center East, School of Medicine, Tokyo Women's Medical University, 1-10, Nishiogu 2-chome, Arakawa-ku, Tokyo 116-8567; [5]GI Oncology, Hospital East, National Cancer Center, Kashiwa-shi, Chiba 277-8577, Japan

Received July 7, 2006; Accepted August 10, 2006

Abstract. Esophageal squamous cell carcinoma (ESCC) shows a high frequency of lymphatic and/or systemic metastasis, even when the tumor invades only the submucosa. To investigate the genetic alterations in circulating esophageal tumor cells, we performed array-based comparative genomic hybridization (CGH) analysis of 8 DNA samples of xenografts, which were previously established from the thoracic duct lymph of 13 ESCC patients. A total of 5 loci (or genes), 10q21.3 (*EGR2*), 11q13.3 (*CCND1/CyclinD1*, *FGF4*, and *EMS1*), 11q14 (*PAK1*), and 22qtel (*ARSA*) were found to be candidate amplified loci in the xenograft. In contrast, a total of 24 loci including 9p21 (*p16* and *MTAP*) were found to be homozygously deleted candidates in the xenograft. Both *p16* homozygous deletion and *CCND1* amplification were detected in 6 (75%) and 5 (62.5%) of the 8 xenografts. Furthermore, by quantitative Southern blot analysis, we found *p16* homozygous deletion in 30.8% (8/26) of the primary tumors and in 50% (4/8) of the metastasized lymph nodes. The frequency of *CCND1* amplification and *p16* homozygous deletion is suggested to be associated with ESCC progression. Matrigel invasion assays of *p16*-deleted ESCC cells showed that restoring wild-type *p16* activity into the cells significantly inhibits tumor-cell invasion, suggesting that *p16* inactivation could be involved in ESCC invasion. This is the first report showing the genetic alteration of concealed tumor cells in the thoracic duct lymph. The present gene list should be helpful for identifying new amplified and deleted genes in primary ESCCs as well as in metastasized lymph nodes.

Correspondence to: Dr H. Sasaki, Genetics Division, National Cancer Center Research Institute, 1-1, Tsukiji 5-Chome, Chuo-ku, Tokyo 104-0045, Japan
E-mail: hksasaki@gan2.res.ncc.go.jp

Key words: array-based comparative genomic hybridization, esophageal cancer, metastasis, thoracic duct lymph, amplification, homozygous deletion

Introduction

In East Asian countries including Japan and China, and in some parts of Europe, esophageal carcinoma consists mainly of squamous cell carcinomas located mostly in the thoracic esophagus, while adenocarcinoma in the distal part of the esophagus has increasingly become the major pathological type found in Europe and North America. Esophageal squamous cell carcinoma (ESCC) is a cancer with one of the poorest prognoses. ESCC shows lymphatic and/or systemic metastasis, even when the tumor invades only the submucosa (1). Therefore, identification of the genetic alterations associated with ESCC progression is thought to be important. However, a comparative study between distantly metastasized tumors and primary tumors is rarely found compared with that between metastasized lymph nodes and primary tumors, because distantly metastasized tumor samples themselves are difficult to obtain. Furthermore, it is quite difficult to identify genetic or epigenetic alterations in circulating tumor cells, since only rare tumor cells exist in the lymphatic duct or blood vessels (2).

Here we performed array-based comparative genomic hybridization (CGH) analysis of DNA samples of the xenografts, which were previously established from the thoracic duct lymph (3), and report that the accumulation of *CCND1* amplification and *p16* homozygous deletion is associated with ESCC progression. Furthermore, matrigel invasion assays of *p16*-deleted ESCC cells showed that restoring wild-type *p16* activity into the cells significantly inhibits tumor-cell invasion.

Materials and methods

Xenografts from thoracic duct lymphs in esophageal cancer. A thoracic duct lymph was collected independently from 13 patients with ESCC during surgery by cannulation into the thoracic duct. The collected volume varied from 20 to 30 ml.

The collected lymph was centrifuged and the pellets were subcutaneously injected into the abdomen of BALB/c-nude mice. Eight established xenografts were previously reported from 8 out of the 13 patients (3). Here we named the 8 xenografts as Xeno-TDL1-8.

Genomic DNA purification from surgical specimens of ESCC patients and xenografts. ESCC tissues were obtained from patients at the National Cancer Center Hospital (Tokyo). Written informed consent was obtained from all the patients. All of the surgical specimens and the 8 xenografts were frozen immediately in liquid nitrogen and stored at -80°C until use. Genomic DNA was extracted from the frozen materials by the conventional phenol-chloroform procedure.

Array-based CGH. The gene copy number was assessed using a commercial array (Genosensor™ Array 300 v1.0, Vysis, IL, USA) according to the manufacturer's protocol. The array contains 287 BAC clones corresponding to various chromosome loci which have been reported to be altered in various human cancers (list available from the manufacturer's web site, http://www.vysis.com/). Briefly, DNA samples isolated from normal human lymphocytes (reference DNA) and tumor samples (test DNA) were labeled by random priming with Cy3- or Cy5-labeled dCTP. The DNA probes (0.1 μg) were mixed with 20 μg of unlabeled Cot-1 DNA and were hybridized to the genomic array, which was then counter-stained with DAPI and analyzed by the fluorescent image capturing system, GenoSensor.

Southern blot analysis. Five micrograms of *Eco*RI-digested DNA per lane was loaded onto 1% agarose gel, and blotted onto a nylon membrane filter, Hybond N$^+$ (Amersham). The probes for the full-length of the *p16* cDNA and the *CCND1* cDNA were labeled with [^{32}P]dCTP using Random Primed DNA labeling kits (Boehringer Mannheim), and hybridized at 42°C in 5X SSC/0.1% sodium dodecyl sulfate (SDS)/50% Dextran for 12 h. The filters were washed three times in 0.1% SSC/0.1%SDS at 65°C, and were exposed to X-ray film at -80°C. To control the contamination of the tumor samples by normal cells, we performed quantitative Southern blot analysis. Hybridization and washing were done under the same stringent conditions as the above procedure. Using a Bio-image-analyzer (BAS2000; Fujix, Kanagawa, Japan), the ratio of the signal intensity of the *p16* gene/a control gene (*PAX-5*) was calculated. Homozygous deletion was defined by the signal intensity of the *p16* gene being <20% of the internal control gene, *PAX-5*, located at chromosome 9p13. For the *PAX-5* probe, a 298-bp DNA fragment was amplified by PCR with the primers (see below) from genomic DNAs.

Genomic PCR amplification of the p16 gene. Sequences of the primers were as follows: A forward primer, 5'-GGTGTT TCTTTAAATGGCTC -3', and a reverse primer, 5'-AGCCT TCATCGAATTAGGTG-3' for *p16*; a forward primer, 5'-GCGGTGCTTCTCCTATGTGAC -3', and a reverse primer, 5'-TTTAAAGTGCTCTGCGTGATG-3' for *PAX-5*. PCR was performed using Takara ExTaq (Takara Corp., Shiga, Japan) in a total volume of 50 μl containing 100 μM of each primer and 50 ng of template DNA. The thermal cycling conditions were 35 cycles of denaturation at 94°C for 1 min, annealing at 55°C for 1 min, and extension at 68°C for 1 min. The last cycle had an additional extension at 68°C for 10 min. The sizes (437 bp of *p16* exon2 and 298 bp of *PAX-5*) and sequences of the PCR products were confirmed by agarose gel electrophoresis and direct sequencing.

Matrigel invasion assay. Two esophageal cancer cell lines, TE1 and TE3, and a mouse fibroblast cell line, STO were used in this study. TE1 has been reported previously to show no alteration of *CCND1* or *p16*, whereas TE3 has shown *p16* homozygous deletion but no *CCND1* amplification (4,5). To assess the infective ability of the adenoviral vectors, the cells were infected with an adenovirus carrying the *E coli* ß-galactosidase gene under the control of the human cyto-megalovirus promoter (Ad-lacZ), and 24 h later they were stained with X-Gal (5-bromo-4-chloro-3-indolyl-ß-D-galactppyranoside). Increased doses of the adenovirus, from 0 to 200 MOI, were used to ascertain the MOI necessary to infect 80% or more of each cell line. The invasion of the eso-phageal tumor cells *in vitro* was measured by the BD BioCoat™ Matrigel™ Invasion Chamber (BD Biosciences) according to the manufacturer's protocol. After infection of Ad-lacZ and an adenovirus carrying *p16* (Ad-*p16*) at 100 MOI, the cells were trypsinized and 500 μm of cell suspension (1x10^6 cells/ml) was added in triplicate wells. After 24-h incubation, the cells that passed through the filter into the lower wells were fixed and stained with 100% methanol and 1% Toluidine blue, respectively. The number of cells invaded was counted by photographing the membrane through a microscope.

Results

Array-based CGH analysis of xenografts derived from thoracic duct lymph of ESCC patients. We previously reported that xenografts were established from the thoracic duct lymph in 8 (61.5%) of the 13 advanced ESCC patients, whereas only 4 (30.8%) patients showed tumor cells in the thoracic duct lymph as revealed by skillful cytologists (3). These facts suggest that circulating tumor cells in the thoracic duct lymph are very few, but have tumor forming activity in nude mice. To conclude this, however, we have to provide more evidence, such as the presence of ESCC-type genetic alterations in the xenograft. The xenografts are composed of mouse mesenchymal cells and human tumor cells. This composition of no contamination of human mesenchymal cells provides an advantage in identifying homozygously deleted loci, which are very difficult to detect by many molecular biological techniques such as genomic subtraction or differential display. To investigate the genetic alterations in this unique material derived from circulating esophageal tumor cells, we performed array-based CGH analysis, which has a great potential for comprehensive analysis of a relative gene-copy number in tumors (6,7) and subjectively enables us to identify new amplified and homo-zygously deleted genes. To investigate the genetic alterations in the xenografts, we used bacterial artificial chromosome (BAC) clone-arrays containing the 287 amplified or lost loci reported previously in each type of tumor (see Materials and

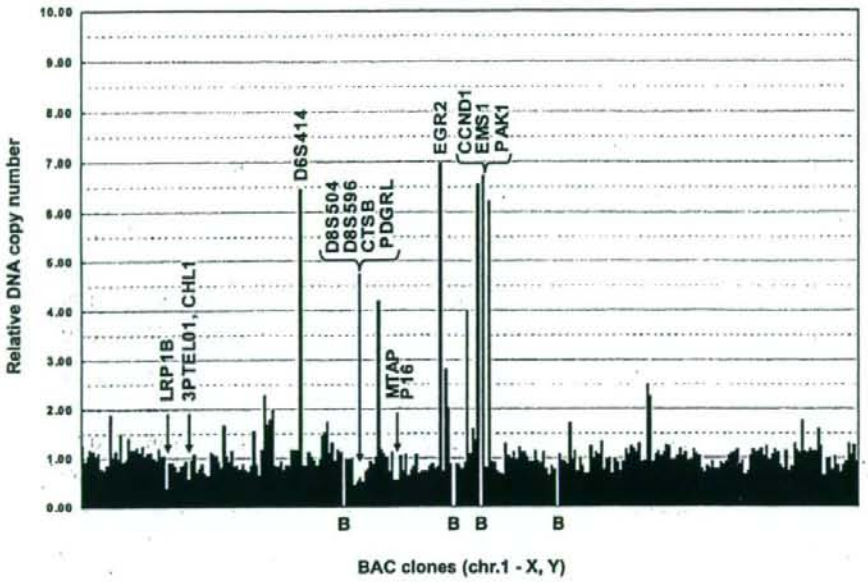

Figure 1. Representative results of array-based CGH on xenograft DNA derived from the thoracic duct lymph of ESCC patients. Fluorescence ratios on all the 287 chromosome loci between a xenograft DNA (Xeno-TDL2 in Table I) from an ESCC patient and a normal male DNA. Amplified or homozygously deleted gene candidates and their chromosome loci, whose ratios between the two samples were changed >5-fold or <0.6-fold (arrows), are also indicated. B, no DNA spot on the array used.

methods). The array-based CGH in the Xeno-TDL2 DNA sample was shown as a representative result (Fig. 1). A >5-fold increased gene (or marker) and its chromosomal locus, which was found in at least one xenograft, are summarized in Table I. A total of 5 loci (or genes), 10q21.3 (*EGR2*), 11q13.3 (*CCND1*, *FGF4*, and *EMS1*), 11q14 (*PAK1*), and 22qtel (*ARSA*) were found to be the candidates amplified in the xenograft. In the same way, a <0.6-fold decreased gene (or marker) and its chromosomal locus are also summarized in Table I. A total of 24 loci were found to be homozygously deleted candidates in the xenograft. Nine telomeric regions, 1qtel (*1QTEL10*), 3ptel (*CHL1*), 4ptel (*SHGC4-207*), 5qtel (*NIB1408*), 8ptel (*D8S504* and *D8S596*), 8q24-qtel (*PTK2*), 12qtel (*stSG8935*), 19ptel (*129F16/SP6* and *stSG42796*), and 19qtel (*D19S238E*) were found to be decreased. The other 15 homozygously deleted candidate loci were 1p12 (*D1S2465* and *D1S3402*), 2q21.2 (*LRP1B*), 3p24.3 (*THRB*), 3p14.2 (*FHIT*), 3q21 (*RBP1*, 2), 3q26.2 (*EIF5A2*), 6q25.1 (*ESR1*), 7q32-34 (*TIF1*), 8p22 (*CTSB*, *PDGRL*, and *LPL*), 9p21 (*p16* and *MTAP*), 9p11.2 (*AFM137XA11*), 10p13 (*BMI1*), 16q24.2 (*CDH13*), 18q11.2 (*LAMA3*), and 18q21.3 (*DCC*), respectively.

CCND1 amplification and p16 homozygous deletion in the xenografts. Among oncogene amplifications in primary ESCCs, *CCND1* amplification has been reported to be most frequent (8). Consistent with these data, *CCND1* amplification was also revealed by array-based CGH analysis of the xenografts (Table I). To confirm the CGH results, we first investigated this gene amplification by a classical but faithful method, Southern blot hybridization, in the 8 xenograft DNA samples. Of them, 5 xenografts (Xeno-TDL1-5) showed *CCND1* amplification (Fig. 2A), thereby providing evidence that the xenograft was derived from the circulating tumor cells in the thoracic duct lymph.

As shown in Table 1, the array-based CGH analysis also showed frequent deletion of the 9p21 locus containing *MTAP* and *p16*. In the 8 xenografts, we next checked for *p16* homozygous deletion by genomic PCR using human specific primers. Six (75%) out of the 8 xenografts showed *p16* homozygous deletion (Fig. 2B). No change in the *p16* copy number in 2 xenografts (Xeno-TDL6 and Xeno-TDL8) shown by genomic PCR was also demonstrated by the array-based CGH (Xeno-TDL6: 1.13 and Xeno-TDL8: 1.17 in Table I). In the xenograft DNA samples, any homozygously deleted genes are detectable by PCR only using human specific primers.

Of the 8 xenografts, only one (Xeno-TDL8) showed no alteration in both *CCND1* and *p16*. Southern blot and genomic PCR analyses of these two genes suggest that most xenografts were derived from the circulating tumor cells in the thoracic duct lymph.

Quantitative Southern blot analysis of p16 in metastasized lymph nodes of ESCCs. We previously reported *p16* mutations in 4 (16%) of 25 primary ESCCs (5). Other investigators successfully detected *p16* homozygous deletion in metastasized lymph nodes (16%, 5/31) by comparative multiplex PCR, and found a decreased amount of *p16* PCR product in 2 out of 5 primary tumors exhibiting *p16* homozygous deletion in metastasized lymph nodes (9). Quantitative PCR analysis provides a quick method for determining the copy number of specific DNA sequences in a large number of clinical samples including paraffin-embedded tissues and biopsy samples.

Table I. Homozygous deleted or amplified candidate loci identified by array-based CGH.

	Gene or marker	Chromosomal loci	Xeno-TDL[a] 1	2	3	4	5	6	7	8
>5-fold	*D6S414*	6p12.1-p21.1	2.27	6.45	1.60	2.58	2.30	1.17	1.27	1.09
	EGR2	10q21.3	6.28	6.94	4.38	6.91	6.45	0.90	1.1	0.78
	CCND1	11q13	5.51	6.53	4.83	6.09	5.88	1.00	1.18	1.07
	EMS1	11q13	5.91	6.71	4.87	6.35	4.58	1.09	1.62	1.3
	PAK1	11q13-q14	5.43	6.20	4.13	6.03	4.55	0.89	1.73	0.93
	9ARSA	22q tel	1.33	1.59	1.01	1.63	1.79	1.10	7.32	1.32
<0.6-fold	*D1S2465, D1S3402*	1p12	0.97	0.94	0.92	0.87	1.07	0.98	0.57	0.87
	1QTEL10	1q tel	0.91	1.25	0.72	1.14	0.89	0.82	0.53	0.82
	LRP1B	2q21.2	0.54	0.39	0.66	0.48	0.56	0.66	0.94	0.5
	3PTEL01, CHL1	3p tel	0.66	0.58	0.70	0.64	0.68	0.39	0.91	0.48
	THRB	3p24.3	0.85	0.71	0.91	0.77	0.84	0.64	0.51	0.66
	FHIT	3p14.2	0.76	0.70	0.83	0.71	0.69	0.47	0.78	0.54
	RBP1, RBP2	3q21-q22	1.17	1.07	1.13	1.04	0.95	1.48	0.54	1.59
	EIF5A2	3q26.2	0.96	0.92	0.92	0.95	0.83	1.46	0.57	1.38
	SHGC4-207	4p tel	0.93	0.83	0.96	0.89	0.81	0.92	0.45	0.88
	NIB1408	5q tel	0.73	0.85	0.71	0.90	0.77	1.04	0.52	1.08
	ESR1	6q25.1	0.84	0.85	0.82	0.87	0.85	0.57	0.85	0.64
	TIF1	7q32-q34	1.15	0.99	1.09	1.00	0.96	1.04	0.57	1.01
	D8S504	8p tel	0.57	0.46	0.59	0.54	0.56	0.53	0.75	0.57
	D8S596	8p tel	0.60	0.52	0.80	0.47	0.75	0.58	0.7	0.54
	CTSB	8p22	0.66	0.60	0.69	0.53	0.62	0.66	1.78	0.47
	PDGRL	8p22-p21.3	0.51	0.52	0.66	0.00	0.76	0.62	1.99	0.8
	LPL	8p22	0.74	0.75	0.81	0.80	0.73	0.65	0.51	0.66
	PTK2	8q24-qter	0.97	1.16	0.96	1.07	1.04	1.47	0.55	1.43
	MTAP	9p21	0.62	0.55	0.79	0.58	0.69	1.08	0.65	1.11
	CDKN2A (p16)	9p21	0.65	0.54	0.80	0.62	0.76	1.13	0.71	1.17
	AFM137XA11	9p11.2	1.15	1.04	1.11	0.95	1.04	0.79	0.38	0.74
	BMI1	10p13	0.89	0.79	1.04	0.81	0.77	1.05	0.51	0.87
	stSG8935	12q tel	1.16	1.15	1.04	1.12	1.11	1.13	0.55	1.13
	CDH13	16q24.2-q24.3	0.74	0.70	0.84	0.77	0.81	0.98	0.54	0.9
	LAMA3	18q11.2	0.69	0.71	0.85	0.72	0.70	0.55	2.16	0.51
	DCC	18q21.3	0.67	0.79	1.12	0.60	0.64	0.81	1.44	0.99
	stSG42796	19p tel	0.99	1.18	0.86	1.20	0.75	1.11	0.44	1.27
	2D19S238E	19q tel	0.67	0.66	0.78	0.77	0.60	0.70	0.61	0.72

[a]Xenografts established from the thoracic duct lymph of ESCC patients.

However, the PCR method is so unstable that we often suffer low reproducibility, and an experiment requires several repetitions (10).

In this study, to examine the frequency of *p16* homozygous deletion in metastasized lymph nodes and primary ESCCs, quantitative Southern blot analysis was performed. Each blot contains 1, 3 and 9 μg of *Eco*RI-digested DNA of normal portions, primary tumors and metastasized lymph nodes to control for possible contamination of the tumor samples by various amounts of normal cells. Representative results of the quantitative Southern blot analysis are shown in Fig. 3. Consistent with previous reports (10,11), a homozygous deletion was defined if the *p16* signal was <20% of the signal from a control gene, *PAX-5*, located on chromosome 9q. We found that *p16* homozygous deletion in primary ESCC and metastatic lymph nodes was detected in 30.8% (8/26) and 50% (4/8) of the cases, respectively (Fig. 3). In summary, *p16* homozygous deletion frequency is likely found to

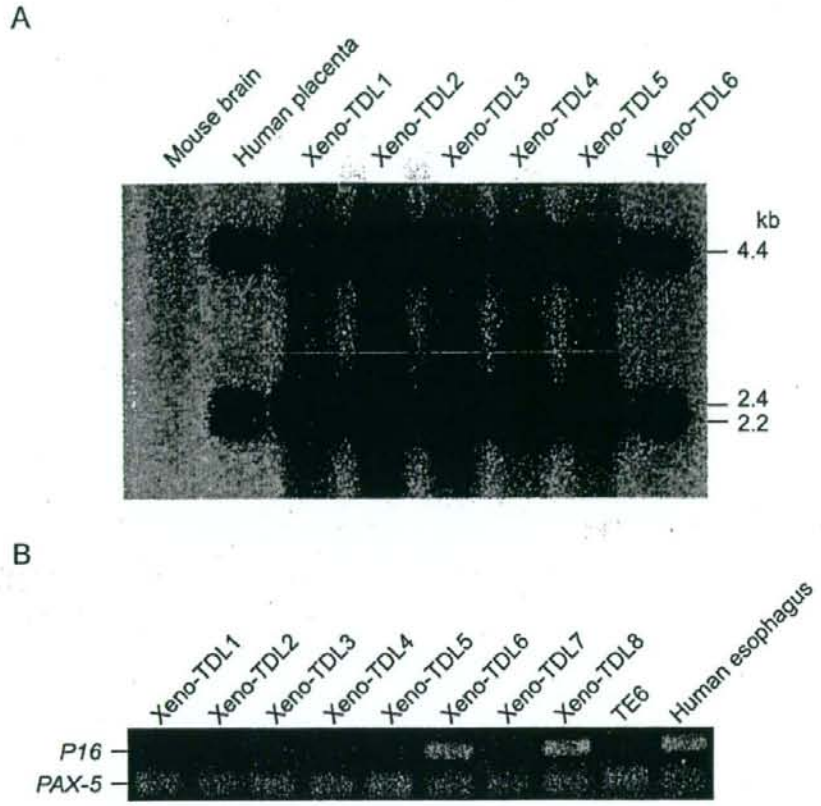

Figure 2. *CCND1* amplification and *p16* homozygous deletion in the xenografts. (A) Southern blot analysis with *CCND1* of 6 xenografts, Xeno-TDL1-6, mouse genome DNA, and human genome DNA. *CCND1* amplification was found in the Xeno-TDL1-5. (B) Genomic PCR of *p16* exon2 and *PAX-5* in 8 xenografts, Xeno-TDL1-8. Two DNA fragments (437 bp of *p16* exon2 and 298 bp of *PAX-5*) amplified by PCR from 50 ng xenograft DNA was analyzed by ethidium bromide-stained 2% agarose gels. An esophageal cancer cell line TE6, in which *p16* has been reported to be deleted, is used as a negative control, and human normal esophagus DNA as a positive control. *p16* homozygous deletion was found in 6 (75%) of the 8 xenografts.

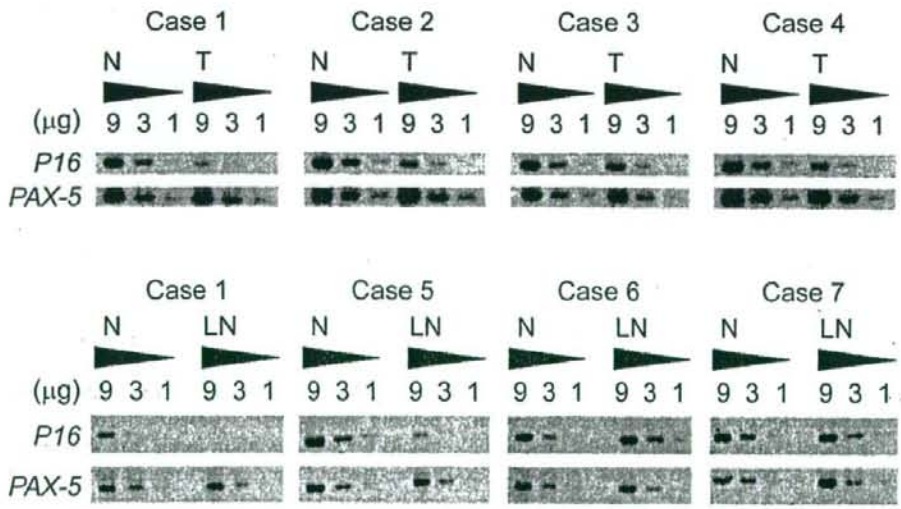

Figure 3. Quantitative Southern blot analysis of the *p16* gene in both primary ESCCs and metastasized lymph nodes. Various amounts of *Eco*RI-digested genomic DNA (9, 3, 1 μg) are loaded to compare the intensity among primary tumor (T), metastasized lymph node (LN) and normal tissue (N). In cases 1, 2, 4, and 5, DNA from the primary tumor or the metastasized lymph node show a remarkable decrease in the signal intensity of *p16* compared to normal tissues, whereas the internal control gene *PAX-5* demonstrated the same intensity in each volume of the genomic DNA between tumor and normal tissues.

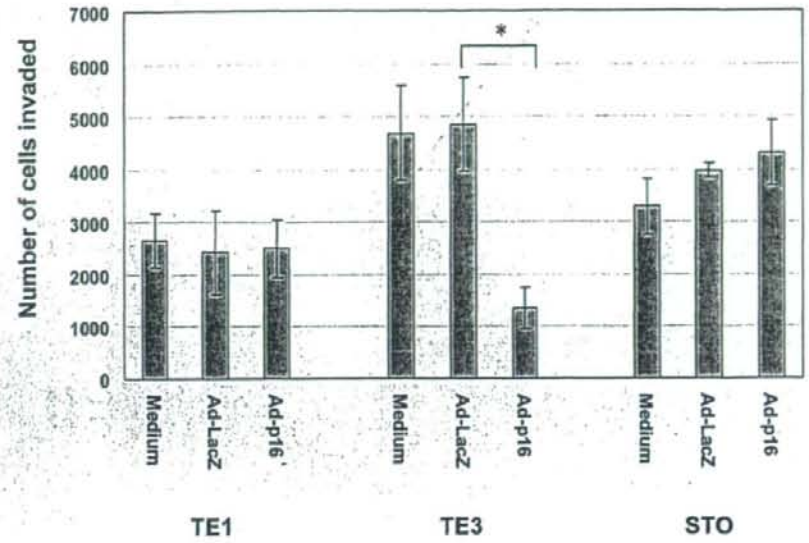

Figure 4. Matrigel invasion assays of esophageal cancer cells infected with Ad-*p16*. The invasion of *p16*-transfected TE3 cells was reduced compared to that of the Ad-lacZ adenovirus control and mock-infected cells; however, no difference of invasion was observed between *p16*-transfected TE1 cells and these two controls. TE1 and TE3 (*p16*-null), esophageal cancer cell lines, and STO, a mouse fibroblast cell line, which is used for a control of invasion assay. *p <0.005.

increase in association with ESCC progression (primary tumors, 30.8%; metastatic lymph nodes, 50%; and circulating tumor cells, 75%).

Adenovirus-mediated p16 gene transfer suppresses invasion of p16-deleted esophageal tumor cells in vitro. The increment of *p16* deletion frequency associated with ESCC progression suggested different functions for *p16* aside from its control of the cell cycle. Therefore, we performed matrigel invasion assays to understand the biological consequences of the *p16* inactivation in ESCC progression. In this assay, we used two esophageal tumor cell lines, TE1 and TE3. TE1 has been reported previously to show no alteration of both *CCND1* and *p16*, whereas TE3 has shown *p16* homozygous deletion but no *CCND1* amplification (4,5). After infection of Ad-lacZ and an adenovirus carrying *p16* (Ad-*p16*) at 100 MOI, which were necessary to infect 80% or more of each cell line (data not shown), the cells were trypsinized and used for the matrigel assay. The invasion of the *p16*-transfected TE3 cells was reduced compared to that of the Ad-lacZ adenovirus control and mock-infected cells; however, no difference of invasion was observed between *p16*-transfected TE1 cells and these two controls (Fig. 4). These results suggest that *p16* inactivation could be involved in ESCC invasion.

Discussion

The amplification frequency (62.5%) of the *CCND1* gene in the xenografts was much higher than that reported previously (28% and 38%) in both 32 primary ESCCs and 13 ESCC cell lines reported previously (8). In regard to primary ESCCs, we previously reported that the 1p34 locus containing *MYCL1*, 2p24 (*MYCN*), 7p12 (*EGFR*), 8p11 (*FGFR1*), and 12q14 (*MDM2*) were amplified in one of the 32 cases (3%), and the 17q12 locus (*ERBB2*) in 2 of the 32 cases (6%), while only the 11q13 locus (*CCND1*, *FGF4*, and *EMS1*) was frequently amplified (28%, 9/32) (8). Another group reported that the 11q22 locus containing *cIAP1* and *MMPs* has been reported to be amplified in 4 of 42 primary ESCCs (9.5%) (12). Therefore, it has been concluded that the 11q13 locus is the most frequently amplified and a major target in ESCC development. *EMS1* in the same amplified locus is known to be involved in invasion and metastasis (13), a function that may account for a report that amplification of the 11q13 locus is useful for predicting outcome and distant organ metastasis in ESCC patients (14).

We found that the *p16* deletion frequency increases in association with ESCC progression (primary tumors, 30.8%; metastatic lymph nodes, 50%; and circulating tumor cells, 75%). Matrigel invasion assays of *p16*-deleted ESCC cells showed that restoring wild-type *p16* activity into the cells significantly inhibits tumor-cell invasion, suggesting that *p16* inactivation could be involved in ESCC invasion. Recently, there is accumulating evidence showing different functions including migration, angiogenesis, and skeletogenesis for *p16* aside from its control of the cell cycle (15,16). It has been reported that adenovirus-mediated *p16* gene transfer suppresses glioma invasion (17). This report also showed that exogenous *p16* expression significantly reduced the expression of matrix metalloproteinase-2 (*MMP-2*), an enzyme involved in tumor-cell invasion. Recently, it has also been reported that *p16* inhibits *MMP-2* expression through the attenuation of Sp1 binding to the *MMP-2* promoter (18). In ESCCs also, the targets for a transcription factor Sp1 should be identified for understanding the detailed mechanism of *p16* in invasion inhibition and for developing new anti-tumor drugs.

Our established xenografts can provide highly sensitive results in detecting gene amplification and deletion by array-based CGH. Many genetic alterations in ESCCs have also been found in other squamous cell carcinomas, especially in

head and neck SCCs. Therefore, the present gene list should be helpful for identifying new amplified and deleted genes in primary tumors as well as in metastasized lymph nodes not only in ESCCs but also in head and neck SCCs.

Acknowledgements

This study was supported in part by a Grant-in-Aid for Cancer Research (16-15) and for the Third Comprehensive 10-Year Strategy for Cancer Control from the Ministry of Health and Welfare of Japan, in part by the Promotion of Fundamental Studies in Health Sciences of the National Institute of Biomedical Innovation (NiBio), and in part by a Research Grant of the Princess Takamatsu Cancer Research Fund. D.-H. Kim is an awardee of Research Resident Fellowships from the Foundation for Promotion of Cancer Research.

References

1. Fahn H-J, Wang L-S, Huang B-S, Huang M-H and Chien K-Y: Tumor recurrence in long-term survivors after treatment of carcinoma of the esophagus. Ann Thorac Surg 57: 677-681, 1994.
2. Zuidema GD: Embryology, anatomy, and physiology of the esophagus. In: Shackelford's surgery of the alimentary tract. 3rd edition. W.B. Saunders Company, Philadelphia, pp27-29, 1995.
3. Kato H, Tachimori Y, Watanabe H, Iizuka T, Itabashi M and Hirota T: Lymph node metastasis in thoracic esophageal carcinoma. J Surg Oncol 48: 106-111, 1991.
4. Igaki H, Sasaki H, Kishi T, Sakamoto H, Tachimori Y, Kato H, Watanabe H, Sugimura T and Terada M: Highly frequent homozygous deletion of the p16 gene in esophageal cancer cell lines . Biochem Biophys Res Commun 203: 3421-3423, 1994.
5. Igaki H, Sasaki H, Tachimori Y, Kato H, Watanabe H, Kimura T, Harada Y, Sugimura T and Terada M: Mutation frequency of the p16/CDKN2 gene in primary cancer in the upper digestive tract. Cancer Res 203: 3421-3423, 1995.
6. Pinkel D, Segraves R, Sudar D, Clark S, Poole I, Kowbel D, Collins C, Kuo W, Chen C, Zhai Y, Dairkee SH, Liung B, Gray JW and Albertson DG: High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays. Nat Genet 20: 207-211, 1998.
7. Pollack JR, Perou CM, Alizadeh AA, Eisen MB, Pergamenschikov A, Williams CE, Jeffrey SS, Botstein D and Brown PO: Genome-wide analysis of DNA copy-number changes using cDNA microarray. Nat Genet 23: 41-46, 1999.
8. Ishizuka T, Tanabe C, Sakamoto H, Aoyagi K, Maekawa M, Matsukura N, Tokunaga A, Tajiri T, Yoshida T, Terada M and Sasaki H: Gene amplification profiling of esophageal squamous cell carcinomas by DNA array CGH. Biochem Biophys Res Commun 296: 152-155, 2002.
9. Maesawa C, Tamura G, Nishizuka S, Ogasawara S, Ishida K, Terashima M, Sakata K, Sato N, Saito K and Satodate R: Inactivation of CDKN2 gene by homozygous deletion and *de novo* methylation is associated with advanced stage esophageal squamous cell carcinoma. Cancer Res 56: 3875-3878, 1996.
10. Walker DG, Duan W, Popovic EA, Kaye AH, Tomlinson FH and Lavin M: Homozygous deletions of the multiple tumor suppressor gene 1 in the progression of human astrocytomas. Cancer Res 55: 20-23, 1995.
11. Tsutsumi M, Tsai YC, Gonzalgo ML, Nichols PW and Jones PA: Early acquisition of homozygous deletions of p16/p19 during squamous cell carcinogenesis and genetic mosaicism in bladder cancer. Oncogene 17: 3021-3027, 1998.
12. Imoto I, Yang ZO, Pimkhaokham A, Tsuda H, Shimada Y, Imamura M, Ohki M and Inazawa J: Identification of *cIAP1* as a candidate target gene within an amplicon at 11q22 in esophageal squamous cell carcinomas. Cancer Res 61: 6629-6634, 2001.
13. Patel AS, Schechter GL, Wasilenko WJ and Somers KD: Overexpression of EMS1/cortactin in NIH3T3 fibroblasts causes increased cell motility and invasion *in vitro*. Oncogene 16: 3227-3232, 1998.
14. Shiozaki H, Ozawa S, Ando N, Tsuruta H, Terada M, Ueda M and Kitajima M: Cyclin D1 amplification as a new predictive classification for squamous cell carcinoma of the esophagus, adding gene information. Clin Cancer Res 2: 1153-1161, 1996.
15. Alhaja E, Adan J, Pagan R, Mitjan F, Cascallo M, Rodriguez M, Noe V, Ciudad CJ, Mazo A, Vilaro S and Piulats J: Anti-migratory and anti-angiogenic effect of p16: a novel localization at membrane ruffles and lamellipodia in endothelial cells. Angiogenesis 7: 323-333, 2004.
16. Cheers MS and Ettensohn CA: P16 is an essential regulator of skeletogenesis in the sea urchin embryo. Dev Biol 283: 384-396, 2005.
17. Chintala SK, Fueyo J, Gomez-Manzano C, Venkaiah B, Bjerkvig R, Yung WK, Sawaya R, Kyritsis AP and Rao JS: Adenovirus-mediated p16/CDKN2 gene transfer suppresses glioma invasion *in vitro*. Oncogene 15: 2049-2057, 1997.
18. Wang CH, Chang HC and Hung WC: P16 inhibits matrix metalloproteinase-2 expression via suppression of Sp1-mediated gene transcription. J Cell Physiol (in press).

Diffuse large B-cell lymphoma after transformation from low-grade follicular lymphoma: morphological, immunohistochemical, and FISH analyses

Akiko Miyagi Maeshima,[1,4] Mutsuko Omatsu,[1] Junko Nomoto,[2] Dai Maruyama,[2] Sung-Won Kim,[2] Takashi Watanabe,[2] Yukio Kobayashi,[2] Kensei Tobinai[2] and Yoshihiro Matsuno[3]

[1]Clinical Laboratory and [2]Hematology and Stem Cell Transplantation Divisions, National Cancer Center Hospital, Tsukiji 5-1-1, Chuo-ku, Tokyo 104-0045, and [3]Department of Surgical Pathology, Hokkaido University Hospital, Kita 14 Nishi 5, Kita-ku, Sapporo 060-8648, Japan

(Received March 12, 2008/Revised April 27, 2008/Accepted May 6, 2008/Online publication June 28, 2008)

Follicular lymphoma (FL) is one of the most common subtypes of non-Hodgkin lymphoma and frequently transforms to diffuse large B-cell lymphoma (DLBCL). To clarify some aspects of the natural history of FL, we retrospectively examined 43 consecutive patients who had DLBCL with pre- or coexisting FL grade 1 or 2. The patients comprised 22 men and 21 women with a median age of 53 years. Most of the patients (34/43) showed advanced-stage (III or IV) disease initially. We examined both FL and DLBCL components morphologically, immunohistochemically, and by interface fluorescence *in situ* hybridization (FISH: IGH/BCL2 fusion, *BCL6 translocation*) analysis. Most of the DLBCLs were classified as the centroblastic subtype, with two exceptions of the anaplastic subtype. Immunohistochemical analysis of both the FL and DLBCL components revealed the following respective positivity rates: CD20 100%/100%, CD10 86%/66%, Bcl-2 96%/91%, Bcl-6 84%/88%, MUM1 16%/34%, CD30 0%/20%, CD138 0%/0%, and CD5 0%/3%. Loss of CD10 (*6/36, 17%*) and gain of MUM1 (*7/28, 25%*) and CD30 (*5/21, 24%*) through transformation were not infrequent. High positivity rates for Bcl-2 and Bcl-6 were maintained throughout transformation. Among the DLBCLs, 84% were classified as the germinal center B-cell phenotype (GCB) and 16% as non-GCB in accordance with the criteria of Hans *et al.* IGH/BCL2 fusion was detected by FISH in 89% of FLs and 82% of DLBCLs. BCL6 translocation was detected in 1/6 (17%) DLBCLs without IGH/BCL2 fusion. Thus, although the morphological features and FISH results for DLBCL were consistent with transformed FL, the immunophenotype showed wide heterogeneity. (*Cancer Sci* 2008; 99: 1760–1768)

Follicular lymphoma (FL) is one of the most common subtypes of non-Hodgkin lymphoma in the Western world, accounting for 22% of all cases worldwide.[1] FL occurs over a broad range of ages, and most cases are manifested initially in lymph nodes.

The risk of FL transformation has been reported as being approximately 20% at 8 years.[2,3] Transformation to DLBCL is observed frequently, with cells most commonly resembling centroblasts,[4] but occasionally resembling anaplastic large cells with CD30 expression.[5] Rare cases have transformed to Burkitt or Burkitt-like lymphoma,[6] or precursor B-lymphoblastic lymphoma/acute lymphoid leukemia.[7] Moreover, composite FL and Hodgkin lymphoma have been suggested to represent two morphologic manifestations of the same tumor clones.[8,9]

In recent years, several analyses of genetic alterations that appear to affect the risk for FL transformation have been reported, including c-MYC translocation,[6] p53 mutation,[10,11] deletions of the tumor suppressor genes p15 and p16,[12,13] and chromosomal 6q23-26 and 17q aberrations.[14] However, there have been few immunohistochemical analyses of transformed FL. FL is positive for the pan-B-cell marker CD20, and frequently positive for CD10, Bcl-2 and Bcl-6, but is usually negative for the postgerminal center B-cell or plasma cell markers CD30, MUM1 and CD138, and CD5. Because only one study has demonstrated gain of CD30 expression through FL transformation,[5] we considered that more analyses were needed to clarify the immunophenotypic changes occurring during the transformation of FL to DLBCL.

Since 2000 DLBCL has been subdivided into GCB and non-GCB (including the activated B-cell phenotype [ABC] and type 3 phenotype) using the cDNA microarray technique.[15,16] The GCB group shows better outcomes and includes cases with translocation (14;18)(q32;q21). For clinical practice, Hans *et al.* showed that a panel of immunohistochemical markers comprising CD10, Bcl-6, and MUM1 could be used on paraffin-embedded tissues to separate DLBCL into tumors with a GCB or non-GCB phenotype.[17] Davies *et al.* examined 35 cases of transformed FL, and found that 89% of them had a GCB phenotype and 9% had a non-GCB phenotype.[18]

The translocation (14;18)(q32;q21) is present in 80–100% of FLs in Western countries,[19,20] whereas in South-East Asia, including Japan, the incidence of translocation is considerably lower: about 60%.[21] It is unclear whether FL with t(14;18) frequently transforms to DLBCL.

The aim of this study was to clarify the natural history of FL, mainly in view of immunophenotypic changes through transformations. We evaluated low-grade FLs and their transformant DLBCLs using morphological, immunhistochemical, and FISH analyses to delineate the heterogeneity of DLBCL after transformation from low-grade FL.

Materials and Methods

Patients. The criteria used for identification of transformed FL were those reported for aggressive B-cell lymphomas in a review based on the workshop of the XIth Meeting of the European Association for Haematopathology.[22] Briefly, it was considered that the term 'transformation' should be used only when there was morphological evidence of simultaneous or prior low-grade FL. Therefore, in the present study, we chose DLBCL with simultaneous or prior low-grade (grade 1 or 2) FL. We retrospectively studied 43 consecutive patients with DLBCL with pre- (20 cases) or coexisting (23 cases) FL grade 1 or 2 treated at the National Cancer Center Hospital, Tokyo, Japan, between 1997 and 2005. The total number of DLBCL specimens was 47 (1–3 per case), and the total number of specimens of low-grade FL was 53 (1–5 per case). A total of 400 FLs and 653 DLBCLs were registered during the same

*To whom correspondence should be addressed. E-mail: akmaeshi@ncc.go.jp

doi: 10.1111/j.1349-7006.2008.00873.x

period. Clinical information was extracted from the medical records, and is summarized in Table 1.

Morphological review. The materials were fixed in 10% neutral-buffered formalin, embedded in paraffin, cut into 4-μm thick sections, and stained with hematoxylin–eosin (HE) for histologic evaluation. All specimens were reviewed by two pathologists (AMM and YM) to confirm that the morphologic characteristics fulfillled the criteria for FL and DLBCL in the 2001 World Health Organization classification of lymphoid neoplasms.[23] Tumors were judged to be FL grade 1 when neoplastic follicles contained 0–5 centroblasts/10 HPF, and FL grade 2 when they contained 6–15 centroblasts/10 HPF. We diagnosed DLBCL when the tumor cells were spread diffusely without a follicular pattern and large lymphoid cells accounted for more than 30% of the tumor cells. DLBCL was subclassified as the centroblastic, anaplastic, immunoblastic, or T-cell/histiocyte rich variant. The centroblastic variant was subclassified as monomorphous (comprising only large lymphoid cells) or polymorphous (comprising a mixture of large- and medium-sized lymphoid cells).

Immunohistochemistry and *in situ* hybridization. We performed immunohistochemical staining for both FL and DLBCL components on formalin-fixed paraffin-embedded tissues using a panel of monoclonal and polyclonal antibodies. Sections 4-μm thick were cut from each paraffin block, deparaffinized, and incubated at 121°C in pH 6.0 citrate buffer for 10 min for antigen retrieval. Antibodies included those against the following antigens: a pan-B-cell marker, CD20 (L26, ×100; Dako, Glostrup, Denmark); a pan-T-cell marker, CD3 (PS1, ×25; Novocastra, Newcastle-upon-Tyne, UK); FL markers, CD10 (56C6, ×50; Novocastra), Bcl-2 (124, ×100; Dako); and Bcl-6 (poly, ×50; Dako, Kyoto, Japan); postgerminal center B-cell or plasma cell markers, CD30 (Ber-H2, ×100; Dako, Denmark); MUM1 (MUM1p, ×50; Dako, Japan); and CD138 (5F7, ×25; Novocastra); and CD5 (4C7, ×50; Novocastra), employing an autostainer with the standard polymer (Dako autostainer plus: CD3, CD5, CD10, and CD30) or labeled streptavidin-biotin method (Biogenex autostainer: CD20 and Bcl-2), or manually by the standard avidin-biotin complex method (Bcl-6, MUM1, and CD138). Immunoreactivity was judged positive if more than 30% of the tumor cells were stained. All immunohistochemical specimens were judged by AM Maeshima, and Y Matsuno confirmed them.

To classify each case as having either a 'GCB phenotype' or a 'non-GCB phenotype', a panel of three antigens (CD10, Bcl-6, MUM1) was used according to the protocol reported by Hans *et al.*[17] Briefly, cases were assigned to the 'GCB phenotype' if the specimens were positive for CD10. If the specimens were negative for both Bcl-6 and CD10, the corresponding cases were assigned to the 'non-GCB phenotype'. If the specimens were positive for Bcl-6 and negative for CD10, the expression of MUM1 was used to determine the group: if MUM1 was negative, the case was assigned to the 'GCB phenotype', and if positive, to the 'non-GCB phenotype'.

Interphase fluorescence *in situ* hybridization (FISH) analysis. Sections 4-μm thick were cut from each paraffin block and used for FISH analysis. The specimens were treated with a 2× saline sodium citrate buffer (SSC, pH 7.3), digested with 0.005% and 0.3% pepsin/0.01 N HCl for 14 min at 37°C, rinsed in 1× phosphate buffer saline (PBS, pH 7.4) for 5 min, formalin $MgCl_2$/PBS for 10 min, rinsed in 1× PBS for 5 min twice, and dehydrated in an ethanol series. Next, the samples were denatured in 70% formamide/20× SSC for 2 min at 37°C and dehydrated with 70% ethanol for 5 min, followed by 100% ethanol. Denatured probes (10 μL) were dropped onto the slides, covered with a coverslip, and sealed with rubber cement. The slides were then treated using a microwave procedure to intensify the signals. The microwave (MI-77; Azumaya Company, Tokyo, Japan) was set to irradiate the samples for 3-second periods at intervals of 2 s, for a total of 60 min at a frequency of 2.45 GHz and an output power of 250 W with the temperature sensor set to 37°C. After incubation overnight at 37°C, the slides were washed with 50% formamide/2× SSC for 10 min at 45°C, and then washed twice more for 10 min each at room temperature; the slides were then washed with 2× SSC for 10 min. The specimens were rinsed in 4× SSC/0.05% Triton for 5 min, 2× SSC for 5 min at 45°C, and 0.2× SSC at room temperature. The slides were covered with antifade solution and viewed under a BX60 fluorescence microscope (Olympus, Tokyo, Japan) using a 100× oil immersion lens and appropriate filters.

LSI IGH Spectrum Green/LSI BCL2 Spectrum Orange Dual Fusion Translocation Probe (Vysis, Downers Grove, IL, USA) was used to detect t(14;18): IGH/BCL2 fusion. LSI BCL6 Dual Color, Break Apart Rearrangement Probe (Vysis) was used to detect 3q27: BCL6 translocation. Judgment of the fusion gene was performed as described previously.[21] Briefly, a total of 50–200 nuclei per case were scored, and if more than 2% of the tumor cells had two fusion signals in the IGH/BCL2 examination, they were judged positive for fusion. If more than 1.5% of tumor cells had split signals in the BCL6 examination, they were judged positive for translocation.

Statistical analysis. Five-year and 10-year overall survival rates were calculated by the Kaplan–Meier method. Univariate analysis was performed using the log-rank test for clinicopathologic parameters, as shown in Tables 1–3. Clinical information of patients with pre-existing FL and those with coexisting FL was compared by the Fisher's exact test, Mann–Whitney *U*-test or log-rank test in Table 2. Differences were considered significant when *P*-value was less than 0.05.

Results

Patients. Clinical information is summarized in Table 1. The patients comprised 22 men and 21 women, ranging in age from 25 to 80 years with a median age of 53 years. Most of them (34/43) had advanced-stage (III or IV) initially. All of the patients received treatments after initial diagnoses: cyclophosphamide, doxorubicin, vincristine and prednisone (CHOP) ± radiation (12 cases), rituximab (R)-CHOP ± radiation (22 cases), or other types of treatment (cyclophosphamide, vincristine, prednisone and procarbazine (C-MOPP), vincristine, cyclophosphamide, prednisone and doxorubicin (VEPA), methotrexate, doxorubicin, cyclophosphamide, vincristine, prednisone and bleomycin (MACOP-B), vincristine/vindesine, doxorubicin and prednisone (VCR/VDS + DOX + PSL), R+ radiation and radiation) (8 cases). Overall survival after initial diagnosis was 76.5% at 5 years and 57.3% at 10 years. Overall survival after transformation was 60.0% at 5 years. Clinical information was compared between 19 patients with pre-existing FL and 24 patients with coexisting FL, and is summarized in Table 2.

Morphology. The results of morphological analysis, immunohistochemical staining for each antibody, and FISH analysis are summarized in Table 3 and Table 4. A total of 47 DLBCL specimens from 43 patients were reviewed. The biopsy sites were lymph node (27), tonsil (6), spleen (1), and extranodal sites (13: gingiva 1, stomach 1, small intestine 1, terminal ileum 1, rectum 3, bone marrow 2, skin 3, lung 1). The subtypes of DLBCL in the final specimens were centroblastic monomorphous (30), centroblastic polymorphous (11), and anaplastic (2). Notably, one case (no. 38) finally transformed to classical Hodgkin lymphoma, mixed cellularity, after transformation of FL to DLBCL. Massive necrosis was detected in four cases (9%).

A total of 53 FL specimens from 43 patients were reviewed. The biopsy sites were lymph node (21), tonsil (2), spleen (1), and extranodal sites (29: nasopharynx 1, esophagus 1, stomach 6, duodenum 6, small intestine 2, colon 1, bone marrow 11, skin 1). The grades of FL were grade 1 (20), grade 2 (27), judged as limited to low-grade FL due to the very small amount

Table 1. Patient characteristics

Case	Age/ gender	Stage	IPI	Initial diagnosis (site)	Initial therapy /response	Interval to transformation in months (time of relapse)	Second diagnosis (site)	Subtype of DLBCL	Follow-up months	Outcome from initial diagnosis
1	50/F	2	L	DLBCL + FL, gr.3a + FL, gr.1 (spleen)	Rx3 + CHOPx7/CR	0		anaplastic	40	AWD
2	47/M	2	L	FL, gr.2 (LN)	CHOPx8/CR	91 (2nd)	DLBCL (tonsil)	centroblastic	96	AWD
3	77/M	3	HI	FL, gr.1 (LN)	VCR/VDS, DOX, PSL/unknown	24 (1st)	DLBCL (LN)	centroblastic	31	DOD
4	45/M	3	L	FL, gr.1 (LN)	C-MOPPx11/CR	71 (4th)	DLBCL + FL, gr.3a (LN)	centroblastic	71	AWD
5	44/M	3	L	FL, gr.2 (LN, BM)	R-CHOPx6/PR	69 (1st)	DLBCL + FL, gr.2 (BM)	centroblastic	69	AWD
6	40/F	4	LI	FL, gr.1 (LN)	R-CHOPx6/CR	38 (2nd)	DLBCL + FL, gr.3a (tonsil), DLBCL (rectum), DLBCL (small intestine)	centroblastic	69	AWD
7	66/F	4	LI	FL, gr.1 (LN)	CHOPx3/PR	11 (1st)	DLBCL + FL, gr.3a (LN), DLBCL (skin)	centroblastic	66	AWD
8	80/F	3	HI	DLBCL + FL, gr.3a + FL, gr.2 (LN)	CHOPx5/PR	0		centroblastic	46	DOD
9	51/F	1	L	FL, gr.2 (duodenum)	radiation 40Gy/CR	10 (1st)	DLBCL (LN)	centroblastic	71	AWOD
10	67/M	3	LI	FL, gr.2 (LN)	VEPAx14/CR	93 (1st)	DLBCL (tonsil)	centroblastic	101	DOD
11	60/M	4	LI	DLBCL + FL, gr.2 (small intestine)	CHOPx8 + radiation 39Gy/CR	0		centroblastic	59	AWOD
12	25/F	4	LI	FL, gr.1 (duodenum), DLBCL + FL, gr.3a (LN)	Rx4 + CHOPx8/CR	0		centroblastic	69	AWOD
13	53/M	4	HI	DLBCL + FL, gr.2 (skin)	CHOPx8/PR	0		centroblastic	23	AWD
14	49/M	4	LI	FL, gr.2 (BM)	MACOP–8x4/PR	49 (2nd)	DLBCL (skin)	centroblastic	55	DOD
15	61/F	3	HI	DLBCL + FL, gr.2 (LN)	Rx4 + CHOPx8/CR	0		centroblastic	44	AWOD
16	46/M	2	L	DLBCL + FL, gr.3a + FL, gr.2 (LN)	R-CHOPx8 + radiation 40Gy/PD	0		centroblastic	19	DOD
17	66/F	3	HI	DLBCL + FL, gr.3a + FL, gr.2 (LN)	CHOPx2 + radiation 40Gy/CR	0		centroblastic	43	AWOD
18	43/M	2	L	DLBCL + FL, gr.1 (LN)	R-CHOPx6 + radiation 40Gy/CR	0		centroblastic	72	AWOD
19	60/F	3	LI	FL, gr.1 (LN, duodenum)	CHOPx8/CR	40 (1st)	DLBCL (tonsil)	centroblastic	80	AWOD
20	41/F	4	LI	FL, gr.2 (tonsil), DLBCL (LN)	R-CHOPx8 + radiation 40Gy/CR	0		centroblastic	38	AWOD
21	72/F	2	LI	FL, gr.2 (LN)	R-CHOPx8 + radiation 40Gy/PR	84 (1st)	DLBCL (LN)	centroblastic	116	AWD
22	46/F	4	H	FL, gr.2 (duodenum, stomach), DLBCL (LN)	R-CHOPx8 + radiation 40Gy/CR	0		centroblastic	32	AWD
23	44/M	3	L	DLBCL + FL, gr.2 (LN)	R-CHOPx8/CR	0		centroblastic	30	AWOD
24	53/F	4	LI	FL, gr.2 (LN)	R + radiation 39Gy/CR	29 (1st)	DLBCL (lung)	centroblastic	29	AWD
25	53/F	4	HI	FL, gr.1 (duodenum, stomach, BM, colon, tonsil), DLBCL + FL, gr.3a (tonsil)	R-CHOPx5/CR	0		centroblastic	33	AWD
26	42/F	4	L	DLBCL + FL, gr.3a (LN), FL, gr.1 (BM)	R-CHOPx6/CR	0		centroblastic	24	AWOD
27	54/M	4	L	FL, gr.2 (LN, BM)	CHOPx8/PR	49 (1st)	DLBCL + FL, gr.3a (BM)	centroblastic	49	AWD

doi: 10.1111/j.1349-7006.2008.00873.x

Table 1. *Continued*

Case	Age/ gender	Stage	IPI	Initial diagnosis (site)	Initial therapy /response	Interval to transformation in months (time of relapse)	Second diagnosis (site)	Subtype of DLBCL	Follow-up months	Outcome from initial diagnosis
28	54/M	4	H	DLBCL + FL, gr.3b + FL, gr.2 (LN)	R-CHOPx8/PR	0		centroblastic	16	AWOD
29	47/M	3	LI	FL, gr.1 (BM)	CHOPx6/CR	46 (2nd)	DLBCL + FL, gr.3b (LN)	centroblastic	47	AWOD
30	59/M	3	L	FL, gr.1 (duodenum), DLBCL + FL, gr.3a (LN)	R-CHOPx6/CR	0		centroblastic	20	AWOD
31	52/M	3	L	DLBCL + FL, gr.3a + FL, gr.2 (LN)	R-CHOPx5/CR	0		centroblastic	18	AWOD
32	57/M	4	L	FL, gr.2 (BM)	R-CHOPx6/CR	26 (1st)	DLBCL + FL, gr.3a (LN)	anaplastic	27	AWD
33	47/F	3	L	DLBCL + FL, gr.1 (LN)	CHOPx6/CR	0		centroblastic	38	AWD
34	47/M	4	LI	FL, low grade (BM), DLBCL (LN)	R-CHOPx8/PR	0		centroblastic	13	AWD
35	47/F	4	L	FL, gr.2 (stomach, BM), DLBCL + FL, gr.3a (LN)	R-CHOPx8/CR	0		centroblastic	7	AWOD
36	59/F	1	L	FL, gr.1 (LN)	CHOPx3 + radiation/CR	24 (1st)	DLBCL (LN)	centroblastic	24	AWD
37	54/M	4	L	FL, low grade (BM), DLBCL + FL, gr.3b (LN)	R-CHOPx6/CR	0		centroblastic	12	AWOD
38	61/F	4	H	FL, low grade (esophagus, stomach), DLBCL + FL, gr.3a (LN), DLBCL (terminal ileum, rectum)	R-CHOPx8/NC	0	HL, MC (LN)	centroblastic	28	DOD
39	68/M	1	L	FL, low grade (BM), DLBCL + FL, gr.3a (LN)	CHOPx3 + radiation/CR	0		centroblastic	8	AWOD
40	79/M	1	L	FL, gr.2 (nasopharynx)	radiation 40Gy/CR	54 (2nd)	DLBCL (stomach)	centroblastic	152	DOD
41	70/M	4	HI	FL, gr.1 (stomach)	Rx4 + CHOPx6/NC	37 (1st)	DLBCL (gingival)	centroblastic	133	DOD
42	45/F	4	L	FL, gr.2 (LN)	C-MOPPx7/CR	21 (1st)	DLBCL (rectum)	centroblastic	37	DOD
43	56/F	4	LI	FL, low grade (stomach), DLBCL (tonsil)	unknown	0		centroblastic	0	unknown

AWD, alive with disease; AWOD, alive without disease; CHOP, cyclophosphamide, doxorubicin, vincristine and prednisone; C-MOPP, cyclophosphamide, vincristine, prednisone and procarbazine; CR, complete remission; DLBCL, diffuse large B-cell lymphoma; DOD, dead of disease; FL, follicular lymphoma; gr. grade; H, high; HI, high intermediate; HL, MC, Hodgkin lymphoma, mixed cellularity; IPI, international prognostic index; L, low; LI, low intermediate; MACOP-B, methotrexate, doxorubicin, cyclophosphamide, vincristine, prednisone and bleomycin; NC, no change; PD, progressive disease; PR, partial remission; R, rituximab; VCR/VDS + DOX + PSL, vincristine/vindesine, doxorubicin and prednisone; VEPA, vincristine, cyclophosphamide, prednisone and doxorubicin.

Cancer Sci | September 2008 | vol. 99 | no. 9 | 1763

Table 2. Patients' clinical data at initial diagnosis

	19 cases with pre-existing FL	24 cases with coexisting FL	*P*-value†
Gender (male/female)	11/8	11/13	0.54
Age (median, range)	57 (40–79)	52 (25–80)	0.30
Stage (I, II/III, IV)	5/14	4/20	0.47
LDH (normal/higher than normal)	13/6	14/10	0.54
PS (0/≥1)‡	13/2	13/7	0.24
Extranodal involvement (0/≥1)	13/6	11/13	0.21
B symptom (−/+)‡	14/1	18/2	0.72
Bulky mass (−/+)	18/1	21/3	0.62
IPI (L, LI/HI, H)‡	14/1	14/6	0.10
5-year OS from initial diagnosis	80.2%	65.2%	0.43
5-year OS from transformation	47.7%	65.2%	0.11

†Fisher's exact test, Mann–Whitney *U*-test or log-rank test.
‡data for PS, B symptom, and IPI were not obtained in eight cases.
FL, follicular lymphoma; H, high; HI, high intermediate; IPI, international prognostic index; L, low; LDH, serum lactate dehydrogenase; LI, low intermediate; OS, overall survival; PS, performance status.

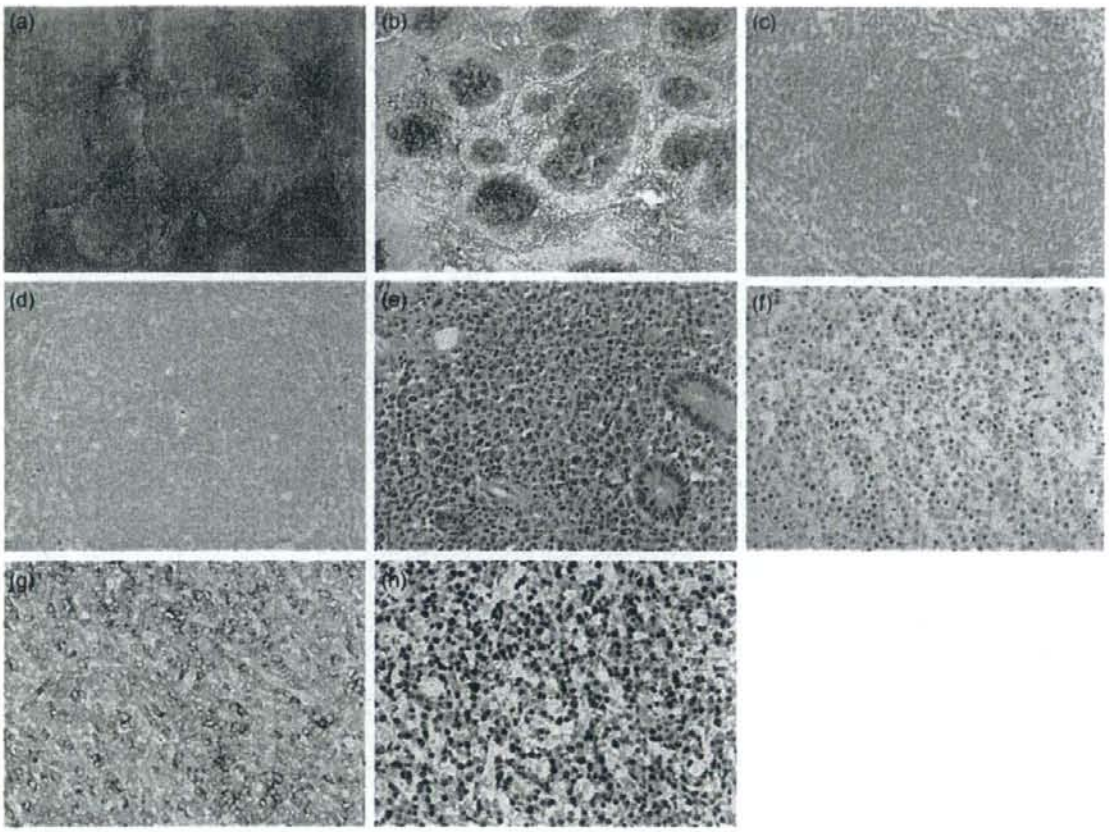

Fig. 1. (a–h) A case of transformation of follicular lymphoma (FL) with a $CD10^+/CD30^-/MUM1^-$ phenotype to diffuse large B-cell lymphoma (DLBCL) with a $CD10^-/CD30^+/MUM1^+$ phenotype. FL, lymph node. (a) Hematoxylin–eosin (HE) (×40), (b) $CD10^+$ (×40), (c) $CD30^-$ (×200), (d) MUM1− (×200). DLBCL, small intestine, (e) HE (×400), (f) $CD10^-$ (×200), (g) $CD30^+$ (×200), (h) $MUM1^+$ (×200).

of material available (6). There were 17 cases with a grade 3a component and three with a grade 3b component.

Immunohistochemistry. The results of immunohistochemistry are summarized in Table 3, Table 4, and Fig. 1. Paraffin sections were available for 41 FL specimens and 43 DLBCL specimens, but in relatively few of the FL and/or DLBCL cases were not available for some of the markers. All tumors were positive for CD20 and negative for CD3. CD10 was positive in 86% (31/36)

doi: 10.1111/j.1349-7006.2008.00873.x

Table 3. Results of immunohistochemistry and FISH analysis

Case no.	Immunophenotype of a low-grade (grade 1 or 2) FL component								FISH[†]		Immunophenotype of a DLBCL component[‡]								FISH		
	CD20	CD10	Bcl-2	Bcl-6	MUM1	CD30	CD5	CD138	IGH/BCL2	BCL6 translocation	CD20	CD10	Bcl-2	Bcl-6	MUM1	CD30	CD5	CD138	GC/non-GC	IGH/BCL2	BCL6 translocation
1	+	+	+	+	−	−	−	−	+		+	+	+	+	−	−	nt	−	GC	+	
2	+	+	+	−	−	nt	−	−	+		+	+	+	+	+	+	+	−	GC	+	
3	+	+	+	+	−	−	nt	−	nt		+	+	+	+	+	−	−	−	GC	+	
4	nt	nt	nt	nt	nt	nt	nt	nt	nt		+	+	+	+	−	−	−	−	GC	+	
5	+	+	+	+	−	−	nt	−	nt		+	nt	+	+	−	−	−	−	GC	nt	
6	+	+	+	+	−	−	−	−	+		+	−	+	−	+	+	−	−	nonGC	+	
7	+	+	+	+	−	−	nt	−	nt		+	−	+	+	+	−	−	−	nonGC	+	
8	+	+	+	+	−	nt	nt	−	+		+	+	+	+	−	−	−	−	GC	+	
9	+	+	+	+	−	−	−	−	+		+	−	nt	nt	nt	nt	nt	nt		+	
10	+	+	+	nt	nt	nt	nt	nt	nt		+	+	−	+	−	−	−	−	GC	+	
11	+	+	+	−	−	−	−	−	nt		+	+	+	−	−	−	−	−	GC	nt	
12	+	+	+	+	+	−	−	−	nt		+	+	+	+	+	−	−	−	GC	−	−
13	+	+	+	+	−	−	nt	−	+		+	+	+	+	−	−	−	−	GC	+	
14	+	nt	+	−	−	−	nt	−	nt		+	+	+	+	−	−	−	−	GC	+	
15	+	+	+	+	−	−	nt	−	+		+	+	+	+	−	nt	−	−	GC	+	
16	+	+	+	+	+	nt	−	nt	+		+	+	+	+	−	nt	−	nt	GC	+	
17	nt	nt	nt	nt	nt	nt	nt	nt	nt		+	+	+	nt	nt	nt	−	nt	GC	nt	
18	+	−	+	+	−	nt	−	nt	nt		+	−	+	+	+	nt	−	nt	nonGC	nt	
19	+	+	+	+	−	−	−	−	nt		+	−	+	nt	nt	nt	nt	nt		nt	
20	+	+	+	+	−	−	nt	−	nt		+	+	+	+	+	−	−	−	GC	+	
21	+	−	−	nt	nt	nt	nt	nt	−	nt	+	−	+	+	−	−	−	−	GC	−	
22	+	+	+	+	−	−	−	−	nt		+	+	+	+	−	−	−	−	GC	+	
23	+	−	+	+	−	−	nt	−	+		+	−	+	nt	nt	nt	−	nt		+	
24	+	+	+	+	−	−	−	−	nt		+	−	+	+	+	+	−	−	nonGC	+	
25	+	+	+	−	+	−	−	−	+		+	+	+	+	+	−	−	−	GC	+	
26	+	+	+	−	−	−	nt	−	+		+	+	+	+	−	nt	nt	nt	GC	+	
27	+	+	+	+	−	nt	nt	nt	+		+	nt	−	+	−	−	−	−	GC	+	
28	+	+	+	nt	nt	nt	nt	nt	+		+	+	+	nt	nt	nt	−	nt	GC	+	
29	+	nt	+	+	+	−	−	−	nt		+	−	+	+	+	+	−	−	nonGC	−	−
30	+	+	+	nt	nt	nt	−	nt	+		+	+	+	+	−	−	−	−	GC	+	
31	+	−	+	nt	nt	nt	nt	nt	nt		+	−	+	nt	−	nt	nt	nt		nt	
32	nt	nt	+	+	+	−	nt	−	nt		+	−	+	nt	+	+	−	nt	nonGC	+	
33	+	+	+	nt	nt	nt	−	nt	nt		+	+	+	nt	nt	−	−	nt	GC	nt	
34	+	+	−,+	+	−	−	−	−	nt		+	+	+	nt	−	nt	−	nt	GC	+	
35	+	+	+	+	−	−	−	−	nt		+	+	+	+	−	−	−	−	GC	+	
36	+	+	+	+	−	−	−	−	nt		+	+	−	+	−	−	−	−	GC	−	−
37	+	nt	+	nt	nt	−	nt	nt	−	nt	+	−	−	nt	nt	nt	−	nt		nt	
38	+	+	+	nt	nt	nt	−	nt	+		+	+	+	−	+	−	−	−	GC	+	
39	+	nt	+	nt	nt	−	nt	nt	nt		+	−	+	+	−	+	−	−	GC	−	+
40	+	+	+	+	−	−	nt	−	nt		+	+	+	−	−	−	−	−	GC	−	−
41	+	+	+	+	−	−	nt	−	nt		+	+	+	+	−	−	−	−	GC	+	
42	+	+	+	+	−	−	−	−	nt		+	−	+	+	−	−	−	nt	GC	nt	
43	+	−	+	nt	nt	nt	−	nt	+		+	+	+	nt	nt	nt	−	nt	GC	+	

[†]judged from the final biopsy specimen, [‡]fluorescence *in situ* hybridization for IGH/BCL2 fusion.
DLBCL, diffuse large B-cell lymphoma; FISH, fluorescence *in situ* hybridization; FL, follicular lymphoma; GC, germinal center B-cell phenotype; nt, not tested.

Table 4. Summary of results of immunohistochemistry and FISH analysis

Antibody	FL component	DLBCL component	Gain/loss/no change
CD20	100% (40/40)	100% (43/43)	
CD10	86% (31/36)*	66% (27/41)†	1/6/29
Bcl-2	96% (42/44)	91% (38/42)	1/4/37
Bcl-6	84% (26/31)	88% (28/32)	4/2/20
MUM1	16% (5/31)	34% (12/35)	7/1/20
CD30	0% (0/28)	20% (6/30)	5/0/16
CD138	0% (0/28)	0% (0/28)	
CD5	0% (0/21)	3% (1/38)	1/0/19
GCB, non-GCB		84% (31/37), 16% (6/37)	
FISH: IGH/BCL2	89% (16/18)	82% (28/34)	

†excluding bone marrow specimens.
DLBCL, diffuse large B-cell lymphoma; FISH, fluorescence *in situ* hybridization; FL, follicular lymphoma; GCB, germinal center B-cell phenotype.

of FLs and 66% (27/41) of DLBCLs, representing a gain in one case, loss in six cases, and no change in 29 cases (including 21 positive cases and eight negative cases). Bone marrow specimens were excluded because only these materials showed extremely low CD10 expression. Among six cases of DLBCL showing loss of CD10, two were diagnosed as DLBCL several times (nos. 6 and 7), and both cases showed loss of CD10 expression between the first and second occasions. Bcl-2 and Bcl-6 were frequently expressed in both FL and DLBCL. Bcl-2 was positive in 96% (42/44) of FLs and 91% (38/42) of DLBCLs, representing a gain in one case, loss in four cases, and no change in 37 cases through transformation. Bcl-6 was positive in 84% (26/31) of FLs and 88% (28/32) of DLBCLs, representing a gain in four cases, loss in two cases, and no change in 20 cases through transformation. Among 32 DLBCL cases for which both Bcl-2 and Bcl-6 immunohistochemistry could be performed, 25 were Bcl-2+/Bcl-6+, four were Bcl-2+/Bcl-6–, 3 were Bcl-2–/Bcl-6+, and no case was Bcl-2–/Bcl-6–.

The postgerminal center B-cell and plasma cell marker MUM 1 was positive in 16% (5/31) of FLs and in 34% (12/35) of DLBCLs, representing a gain in seven cases, loss in one case, and no change in 20 cases (including four positive cases and 16 negative cases). CD30 was negative in all FLs and positive in 20% (6/30) of DLBCLs. Two cases of DLBCL with anaplastic morphology were positive for CD30. CD30 showed scattered expression in marginally and sparsely distributed large lymphoid cells of low-grade FL and FL grade 3, but no case showed positivity in over 30% of the cells. CD5 was negative in all FLs and positive in only one case (no. 2) of DLBCL. This positive case was FL grade 2 in an abdominal lymph node with a $CD10^+$/Bcl-2+/Bcl-6–/$CD5^-$/cyclin D1-immunophenotype and IGH/BCL2 fusion by FISH analysis initially, and showed transformation to centroblastic monomorphous DLBCL in the tonsil, revealing a $CD10^-$/Bcl-2+/Bcl-6+/$CD5^+$/cyclin D1-immunophenotype and IGH/BCL2 fusion by FISH analysis. In the one case of classical Hodgkin lymphoma after transformation from FL via DLBCL (no. 38), FL in the stomach and esophagus and DLBCL in a cervical lymph node had a $CD20^+$/$CD30^-$/$CD10^+$ phenotype and IGH/BCL2 fusion by FISH analysis, but Hodgkin/Reed-Sternberg cells in an inguinal lymph node had a $CD20^-$/$CD30^+$/$CD15^+$/$CD10^-$ phenotype, and were negative for EBER-1 *in situ* hybridization and positive for IGH/BCL2 fusion by FISH (Fig. 2).

Thirty-one (84%) DLBCLs were classified as GCB, and six (16%) DLBCLs were classified as non-GCB. Two cases of DLBCL (nos. 6 and 7) for which several sequential biopsies were taken were judged from the final biopsy specimen: these were non-GCB in the final DLBCL specimens, but had been GCB in the initial specimens.

FISH analysis. Paraffin-embedded sections were available for 18 FL cases and 34 DLBCL cases. IGH/BCL2 fusion was detected in 89% (16/18) of FL cases and in 82% (28/34) of DLBCL cases. In all six DLBCL cases without IGH/BCL2 fusion, BCL6 translocation was detected in one case (17%). Two FL cases without IGH/BCL2 fusion were not available paraffin-embedded sections.

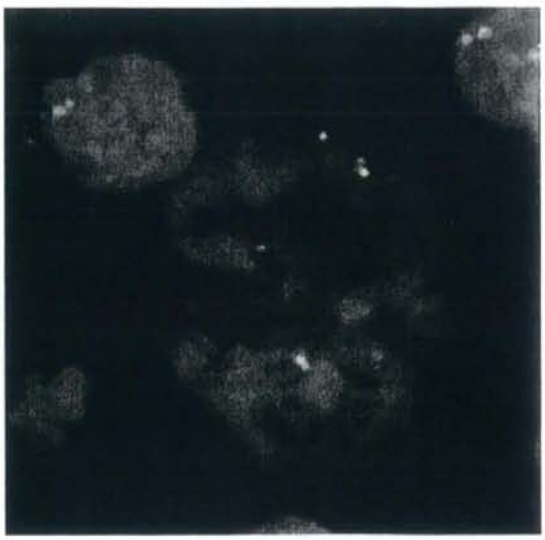

Fig. 2. The result of fluorescence *in situ* hybridization of classical Hodgkin lymphoma transformed from follicular lymphoma (FL). IGH and BCL2 fusion pattern with the LSI IGH Spectrum Green/LSI BCL2 Spectrum Orange Dual Fusion Translocation Probe (Vysis, Downers Grove, IL, USA). Two fusion IGH/BCL2 signals are present.

Statistical analysis. Only the initial treatment regimen was a significant prognostic factor: patients who received CHOP or R-CHOP showed a better outcome than patients who received other treatments ($P < 0.001$). No other significant prognostic factors were detected, including GCB *versus* non-GCB. However, patients with DLBCL showing CD30-positivity (six cases) did not die as a result of disease progression.

Discussion

As transformation of FL to DLBCL is currently the focus of widespread clinical and pathological interests, we studied the

doi: 10.1111/j.1349-7006.2008.00873.x