Multiple fuzzy neural network system for outcome prediction and classification of 220 lymphoma patients on the basis of molecular profiling

Tatsuya Ando,[1] Miyuki Suguro,[2] Takeshi Kobayashi,[1] Masao Seto[2] and Hiroyuki Honda[1, 3]

[1]Department of Biotechnology, School of Engineering, Nagoya University, Furo-cho, Chikusa-ku, Nagoya 464-8603; and [2]Division of Molecular Medicine, Aichi Cancer Center Research Institute, 1-1 Kanokoden, Chikusa-ku, Nagoya 464-8681

(Received June 11, 2003/Revised August 6, 2003/Accepted August 7, 2003)

A fuzzy neural network (FNN) using gene expression profile data can select combinations of genes from thousands of genes, and is applicable to predict outcome for cancer patients after chemotherapy. However, wide clinical heterogeneity reduces the accuracy of prediction. To overcome this problem, we have proposed an FNN system based on majoritarian decision using multiple noninferior models. We used transcriptional profiling data, which were obtained from "Lymphochip" DNA microarrays (http://llmpp.nih.gov/DLBCL), reported by Rosenwald (*N Engl J Med* 2002; 346: 1937–47). When the data were analyzed by our FNN system, accuracy (73.4%) of outcome prediction using only 1 FNN model with 4 genes was higher than that (68.5%) of the Cox model using 17 genes. Higher accuracy (91%) was obtained when an FNN system with 9 noninferior models, consisting of 35 independent genes, was used. The genes selected by the system included genes that are informative in the prognosis of Diffuse large B-cell lymphoma (DLBCL), such as genes showing an expression pattern similar to that of CD10 and BCL-6 or similar to that of IRF-4 and BCL-4. We classified 220 DLBCL patients into 5 groups using the prediction results of 9 FNN models. These groups may correspond to DLBCL subtypes. In group A containing half of the 220 patients, patients with poor outcome were found to satisfy 2 rules, i.e., high expression of MAX dimerization with high expression of unknown A (LC_26146), or high expression of MAX dimerization with low expression of unknown B (LC_33144). The present paper is the first to describe the multiple noninferior FNN modeling system. This system is a powerful tool for predicting outcome and classifying patients, and is applicable to other heterogeneous diseases. (Cancer Sci 2003; 94: 906–913)

Diffuse large B-cell lymphoma (DLBCL) is the largest category of lymphoid malignancy, accounting for 30 to 40% of non-Hodgkin's lymphomas.[1)] From 35 to 40% of DLBCL patients can be cured by current chemotherapeutic regimens; the remaining 60 to 65% eventually succumb to this disease. Marked clinical heterogeneity of DLBCL has been reported by many researchers.[2–5)] Some patients are not correctly diagnosed, and diagnostic categories have not been defined molecularly.[6–10)]

Molecular analyses of clinical heterogeneity in DLBCL have been conducted using microarray technologies. Various computational analyses using small numbers of DLBCL patients have been presented.[11–14)] Previously, we used a fuzzy neural network (FNN) model to identify several genes with which we constructed noninferior prognostic models of DLBCL, and we achieved accuracy greater than 90% using these models.[13, 14)] The FNN model is a relatively advanced artificial neural network (ANN) model. The FNN model has the advantages of extremely high prognostic accuracy and the ability to explicitly describe causality between input and output variables as linguistic IF-THEN rules.[15)] We extracted prognostic rules based on the expression profiles of only 4 genes. In a study on the data of 40 patients reported by Alizadeh *et al*.,[11)] a rule based on 2 prognostic genes (*CD10* and *IRF-4*) was able to identify most of the patients who would later have a poor outcome. Recently, Rosenwald *et al*. analyzed a large number of patients to classify subtypes and predict outcome,[16)] but the predictive value of their model was lower than that of models used for small numbers of patients. This is partly because they only constructed 1 model, despite having analyzed heterogeneous groups.

In the present study, we constructed multiple noninferior FNN models, taking advantage of quick gene selection, to improve prediction accuracy. We then used these models to classify DLBCL patients.

Materials and Methods

Data preprocessing. Transcriptional profiling data obtained from "Lymphochip" DNA microarrays (http://llmpp.nih.gov/DLBCL)[16)] were used in the present study. A total of 7399 gene expression data from 240 DLBCL patients with long-term clinical follow-up were available. Twenty patients were censored, because their follow-up ceased within 4 years from the beginning of anthracycline-based chemotherapy. We divided the remaining 220 patients into 2 groups: those who were alive 4 years after the beginning of anthracycline-based chemotherapy (group 1; n=102), and those who were dead within 4 years from the beginning of anthracycline-based chemotherapy (group 2; n=118). All gene expression data that were not available for more than 10% of these 220 patients were also censored. Thus, 7384 genes spotted on the microarray were used for model construction. The expression ratio of each gene was normalized to the range of 0.1 to 0.9. For outcome classification, patients in group 1 were assigned a value of 0.1, and those in group 2 were assigned a value of 0.9.

Construction of FNN model. A Type-I FNN[17)] was used to establish the relationship between gene expression and clinical outcome. The SWEEP operator method[18)] was used in association with simplified FNN modeling, in which gene selection can be performed quickly. The procedure for modeling using the SWEEP operator method is described briefly below.

In the first step, we predicted the outcome of each patient using the FNN model with expression of only 1 gene, with model parameters optimized using the SWEEP operator method. The prediction model for all 7384 genes was constructed in series, and each gene was ordered according to the accuracy of the constructed model. The 9 genes with the highest accuracy levels were used for the following step to construct 9 noninferior models.

In the second step, we constructed a 2-input model, in which a ranked gene was designated as input and a partner gene was

[3]To whom correspondence should be addressed.
E-mail: honda@nubio.nagoya-u.ac.jp

used to predict outcome using the SWEEP operator method and parameter increasing method (PIM). Repeating this step, we identified a combination of 6 candidate genes for use in model construction, with respect to the input gene. Then, we repeated this procedure for other genes ranked in the first step.

In the final step, we assessed multiple combinations of 6 genes, i.e., from the first-selected to sixth-selected genes. We constructed 6 FNN models, from 1 input using only the first-selected gene to 6 inputs using all 6 genes. The FNN models were strictly constructed using a back-propagation algorithm. Performance of the FNN model was assessed by cross-validation.[19] The learning ratio and learning time were set at 0.1 and 5000 iterations, respectively. In each model, 146 or 147 (=2/3) data were used for learning, and 74 or 73 (=1/3) data were used for evaluation. Output was transferred to the predicted outcome using a threshold value of 0.5: group 2 for values greater than 0.5, and group 1 for less than 0.5. FNN models were assessed according to the accuracy of evaluation data in cross-validation. Accuracy of the model was defined as the ratio of patients that received the correct output. The model with the highest accuracy among the 6 models was selected as the best model and it is a noninferior model.

The discrimination ratio of the model was defined as the ratio of patients within the range of discrimination described by the equation

$$|O_i - T_i| < 0.3 \quad i = 1 \text{ to } 220$$

where O_i and T_i are the output signal of patient i and the teacher signal (0.1 or 0.9) of patient i, respectively.

Majoritarian decision using multiple noninferior modeling system. At the model construction step, 9 noninferior models corresponding to 9 genes with the highest accuracy levels were constructed. The prediction results of each noninferior model were used to improve accuracy. We predicted survival versus death using each model, and determined that a patient would survive if the majority of models predicted they would survive. If a patient had equal numbers of predictions of survival and death, that patient was designated as not discriminable.

Clustering analysis. We classified DLBCL patients based on the accuracy of the prediction of each FNN model. We classified the genes according to expression ratio in DLBCL patients. Hierarchical clustering was applied to this analysis, and the results are shown as a TreeView.[20]

Statistical analysis. The P values of genes were calculated using the t test. The P values for significance of differences between subgroups were calculated using the χ^2 test. The Kaplan-Meier survival analysis plots were calculated using WinSTAT Statistics for Windows Version 3.1 (LightStone Co., Tokyo). The significance of differences in survival rates was analyzed using a log-rank test (Mantel-Cox method).

Results

Ranking of causal genes. Using the SWEEP operator method, we calculated the prediction accuracy of a 1-input FNN model, and ranked 7384 genes using microarray data. The top 15 genes are listed in Table 1. The P values of these genes indicated significance. Some of these genes have previously been reported to be prognostic factors in DLBCL outcome. The genes *EST*, *AA805575* and *CD9* show a correlation with outcome[16, 21], and are overexpressed in cured DLBCL patients. Protein kinase C β1 also shows a correlation with outcome,[12] and is overexpressed in fatal cases of DLBCL. The presence of known prognostic factors among our highest ranked genes indicates that the present method is as suitable for selection of causal genes as a conventional statistical method.

The FNN model using the first-ranked gene. The partner genes of the first-ranked gene, *AA805575*, were selected in order to obtain the best combinations with this gene. The FNN model with 4 genes achieved the highest prediction accuracy. This model is one of the noninferior models. In this model, the genes for sentrin/SUMO-specific protease 3, butyrophilin (subfamily 3, member A2) and tumor necrosis factor (ligand) superfamily (member 10) were selected as partner genes. The earlier the genes were selected, the more effective they were as input for outcome prediction. The accuracy of the FNN model with 4 genes was 73.4%, whereas that of a Cox proportional-hazards model with 17 genes was 68.5%.[16] Hierarchical clustering of those 4 genes based on expression ratios of DLBCL patients was performed. The genes selected by FNN were classified into clusters containing germinal-center B-cell markers (*AA805575*), lymph-node markers (sentrin/SUMO-specific protease 3) and proliferation signatures (butyrophilin). Those types of markers have been reported as prognostic factors. These results indicate that FNN can extract a combination of genes related to outcome from a very large number of genes without prior knowledge.[11, 16]

Optimal number of gene combinations. In the previous study,[13] we constructed an FNN model consisting of 4 genes (*CD10*, *AA807551*, *AA805611* and *IRF-4*) for outcome prediction of 40 DLBCL patients, and this model predicted prognosis with 93% accuracy. For the 58 DLBCL patient data reported by Shipp *et al.*,[12] we constructed FNN models with combinations of 10 genes. Accuracy greater than 80% was achieved with most of these models, and one of these models predicted patient outcome with 93% accuracy.[14] In the present paper, 73.4% accu-

Table 1. Fifteen highest-ranked genes for predicting outcome using SWEEP operator method

Rank	Gene	Accession No.	*P* value
1	ESTs, weakly similar to A47224 thyroxine-binding globulin precursor	AA805575	<0.0001
2	protein tyrosine phosphatase, non-receptor type 2	AA477822	<0.0001
3	ESTs, weakly similar to neuronal thread protein	AA279391	0.0006
4	protein tyrosine phosphatase, non-receptor type 2	AA193262	<0.0001
5	NADH dehydrogenase (ubiquinone) flavoprotein 2 (24 kD)	AA809629	0.0089
6	ESTs	AA714020	0.0482
7	MAD, mothers against decapentaplegic homolog 4	AA441842	0.0058
8	ribosomal protein L13	AI479066	0.0286
9	unknown A		0.0175
10	glutathione synthetase	AA284323	0.0354
11	ESTs, weakly similar to VAMP-5 (MYOBREVIN) (HSPC191)	AI436418	0.0029
12	protein kinase C, β1	AA262174	0.0235
13	regulator of G-protein signalling 16	AA292532	0.0003
14	ribosomal protein S3A	M84711	0.0011
15	CD9 antigen (p24)	AA412053	0.001

racy was obtained using one FNN model. The accuracy of 73.4% was relatively low compared with the accuracy described in the previous papers, although it was higher than that of the Cox proportional-hazards model described above.

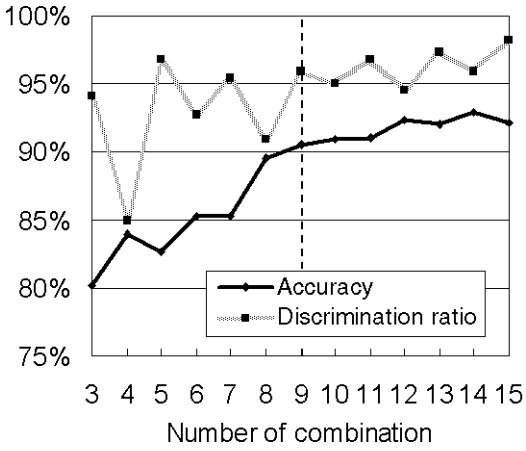


Fig. 1. Accuracy and discrimination ratio by majoritarian decision using multiple noninferior modeling system. Accuracy (bottom line with circle) and discrimination ratio (top line with square) are plotted against the number of combinations of genes.

This lower accuracy than that described in the previous papers may be due to the greater number of patients analyzed in the present study, and the molecular heterogeneity of DLBCL patients. It is unlikely that 1 FNN model with a combination of genes can perfectly predict patient outcome. Assuming that an FNN model can predict outcome of a part of DLBCL patients, integration of various models may help improve prediction accuracy. Accordingly, we selected not only partner genes for the first-ranked gene (*AA805575*), but also partner genes for the lower-ranked genes, i.e., 9 genes with the highest accuracy levels. Multiple noninferior FNN models were constructed using these combinations of genes. From majoritarian decision using the multiple noninferior modeling system, prediction accuracy increased as the number of genes used in the combinations increased (Fig. 1). Majoritarian decision using 9 models produced accuracy greater than 90% and a discrimination ratio greater than 95%. When the system consisting of more than 9 models were employed for prediction, there were no marked increases in accuracy or discrimination ratio. Therefore, we concluded that the present system with 9 models is optimal.

Comparison of selected genes and known prognostic factors. The 9 combinations of genes selected as inputs of FNN are listed in Table 2; the number of independent genes used in these combinations was 35. We compared the expression profiles of selected genes with those of known prognostic factors. The genes for CD10, BCL-2 and BCL-6,[22, 23] were selected as established

Table 2. Nine combinations of genes selected by SWEEP operator method

		Gene	Accession No.	*P* value
1	1	ESTs, weakly similar to A47224 thyroxine-binding globulin precursor	AA805575	<0.001
	2	sentrin/SUMO-specific protease 3	AA742330	0.015
	3	butyrophilin, subfamily 3, member A2	U90143	0.738
	4	tumor necrosis factor (ligand) superfamily, member 10	H54629	0.794
2	1	protein tyrosine phosphatase, non-receptor type 2	AA477822	<0.001
	2	CD81 antigen (target of antiproliferative antibody 1)	H69729	0.018
	3	oxysterol binding protein-like 3	AB014604	0.113
3	1	ESTs, weakly similar to neuronal thread protein	AA279391	<0.001
	2	mastermind-like 1 (*Drosophila*)	AA252661	0.079
	3	ESTs	AA278908	0.621
	4	ESTs, highly similar to EFHU1 translation elongation factor	AA261967	0.852
4	1	protein tyrosine phosphatase, non-receptor type 2	AA193262	<0.001
	2	CD81 antigen (target of antiproliferative antibody 1)	H69729	0.018
	3	MHC class II transactivator	AA287083	0.707
	4	interferon consensus sequence binding protein 1	N62269	0.949
5	1	NADH dehydrogenase (ubiquinone) flavoprotein 2 (24 kD)	AA809629	0.009
	2	E74-like factor 2 (ets domain transcription factor)	AA426604	0.333
	3	ADP-ribosyltransferase (NAD+; poly (ADP-ribose) polymerase)	AA281973	0.152
	4	hypothetical protein	AA243466	0.183
	5	protein phosphatase 1, regulatory (inhibitor) subunit 1A	AA460827	0.041
6	1	ESTs	AA714020	0.048
	2	protein tyrosine phosphatase, receptor type, N polypeptide 2	AA464590	0.364
	3	von Hippel-Lindau syndrome	AA909708	0.690
	4	early B-cell factor	AA279473	0.151
7	1	MAD, mothers against decapentaplegic homolog 4	AA441842	0.006
	2	nuclear receptor coactivator 1	AA180462	0.064
	3	putative chemokine receptor; GTP-binding protein	AA631849	0.804
	4	unknown C		0.011
8	1	ribosomal protein L13	AI479066	0.029
	2	hypothetical protein MGC14859	AI056455	0.048
	3	SH2 domain-containing phosphatase anchor protein 1	AA287337	0.479
	4	CD97 antigen	AA678757	0.575
	5	MAD, mothers against decapentaplegic homolog 4	AA441842	0.022
9	1	unknown A		0.017
	2	nuclear receptor subfamily 3, group C, member 1	AA252235	0.005
	3	MAX dimerization protein	H86558	0.005
	4	unknown B		0.111

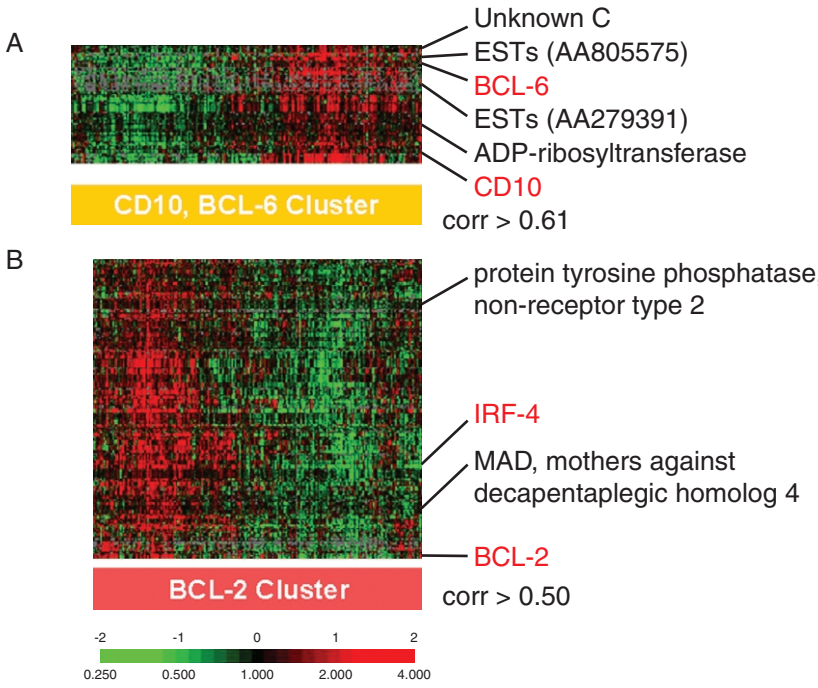


Fig. 2. Gene expression cluster of known prognostic factors defined by hierarchical clustering. Panel A shows the CD10 and BCL-6 cluster including 4 genes found in combinations 1, 3, 5 and 7. Panel B shows the BCL-2 cluster including IRF-4 and 2 genes found in combinations 2, 4, 7 and 8. Each column and row is a patient and a gene, respectively. Expression profiles of 220 patients are shown. These ratios are measures of relative gene expression in each experiment sample, and are depicted according to the color scale shown at the bottom. As indicated, the scale extends from fluorescence ratios of 0.25 to 4 (−2 to +2 log base 2 units). Gray indicates missing or excluded data.

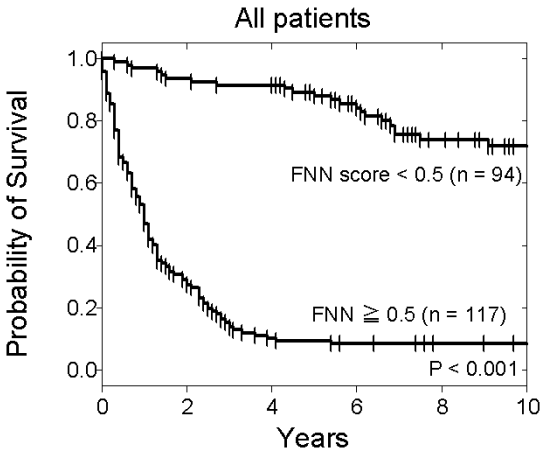


Fig. 3. Kaplan-Meier plot of 4-year overall survival for all patients grouped on the basis of the FNN. The patients predicted to survive on the basis of the FNN models had a 4-year overall survival of 91%. The patients predicted not to survive had a 4-year overall survival of 10%. The tick marks represent censored data.

prognostic factors of DLBCL. We applied a hierarchical-clustering algorithm to the original set of 7384 genes, to check for clusters containing these 3 factors. The factors CD10, BCL-2 and BCL-6 occurred in 2 clusters: a cluster containing CD10 and BCL-2, and a cluster containing BCL-6. Together, these 2 clusters contained 6 genes among 35 genes listed in Table 2. Based on the results of analysis, the 9 combinations of genes selected were divided into 3 groups. The first group included the genes with the same expression pattern as those of *CD10* and *BCL-6*, and consisted of combinations 1, 3, 5 and 7. The second group included genes with the same expression pattern as that of *BCL-2*, and consisted of combinations 2, 4, 7 and 8 (Fig. 2). Combination 7 included 2 genes that occurred in the 2 clusters containing the 3 prognostic factors. The third group consisted of combinations 6 and 9; these combinations did not include genes with the same expression pattern as those of the prognostic factors. As discussed below, combination 7 produced the largest number of accurate predictions, and combination 6 was the best for prediction with patient groups that showed relatively low accuracy when the known signatures were used for prediction.

Kaplan-Meier plot of overall survival of DLBCL patients grouped on the basis of majoritarian decision using multiple noninferior modeling system. Kaplan-Meier survival analysis was performed based on the majoritarian decision of 9 FNN models. The results clearly divided the patients into 2 groups ($P<0.001$, Fig. 3). Patients predicted to survive by the FNN models had significantly longer survival time (4-year overall survival [OS], 91%) than those predicted not to survive (4-year OS, 10%). Furthermore, Kaplan-Meier plots of overall survival showed that the groups based on the FNN model were independent of the groups classified by the International Prognostic Index (IPI).[2)] Among the patients whose IPI classified them in the low-risk group, FNN predicted that 24 patients would have a poor outcome (Fig. 4A), and 17 of these 24 died during the 4-year observation period. Among the patients whose IPI classified them in the intermediate-risk group, FNN predicted that 62 patients would have a poor outcome (Fig. 4B), and 59 of these 62 died during the 4-year observation period. Of the 29 patients classified in the high-risk group by IPI, 23 were predicted not to survive by FNN, and all but 2 of these 23 died during the 4-year observation period (Fig. 4C), and 20 died within 2 years of the prediction. The patients predicted to have poor outcome by both IPI and FNN were found to have significantly high clinical risk.

Subgroup analysis of DLBCL. We subdivided 220 DLBCL patients based on the accuracy of predictions by each FNN model. Hierarchical-clustering was performed using the results of predictions by 8 FNN models. Model 4 was excluded from this clustering analysis, because models 2 and 4 included the same 2 genes, and these 2 FNN models showed similar discrimination for almost the same group of patients. DLBCL patients were subdivided into 5 subgroups (Fig. 5). The largest subgroup, group A, consisted of about half the patients. In group A, 6 of the 8 FNN models had a prediction accuracy greater than 80%, and model 9 had a prediction accuracy of 96.7%. The prediction accuracy of the FNN models was relatively low for the other groups. Consequently, at least one of 8 indepen-

dent FNN models had prediction accuracy greater than 90% in each group. The group with the lowest prediction accuracy was group D. Only model 7 could predict outcome of patients in group D with accuracy greater than 90%.

The subgroups we proposed did not correlate with survival time or risk groups defined on the basis of IPI (Fig. 5). For example, in groups C and D, 4-year survival was relatively low, but a relatively low number of high-risk patients was identified

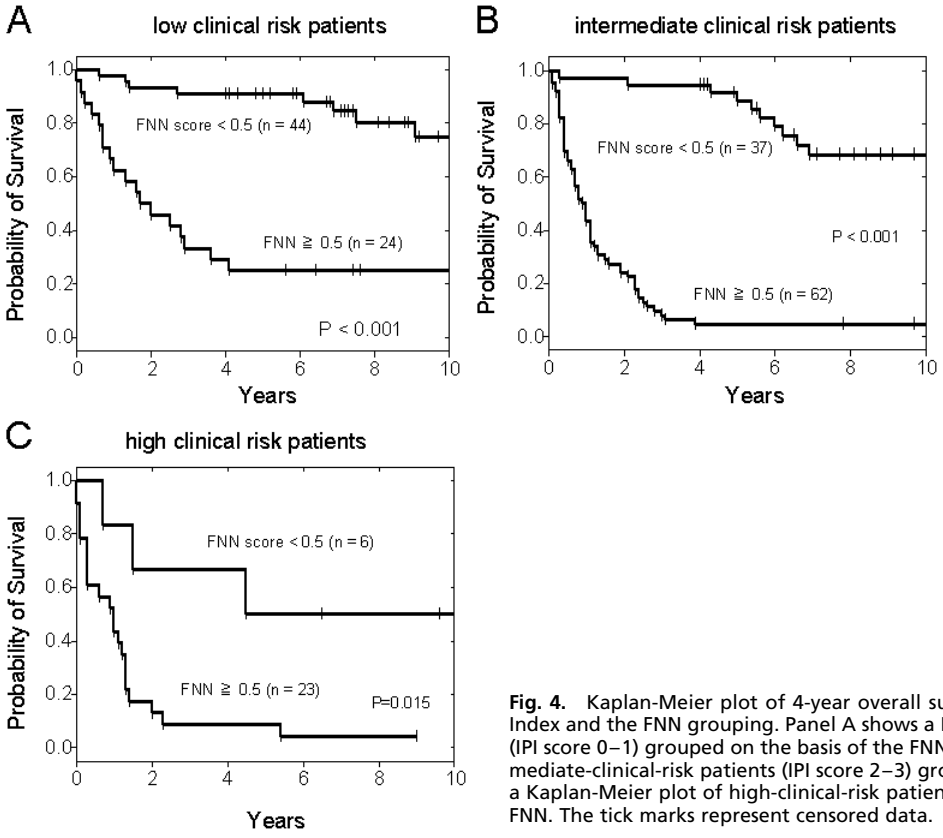
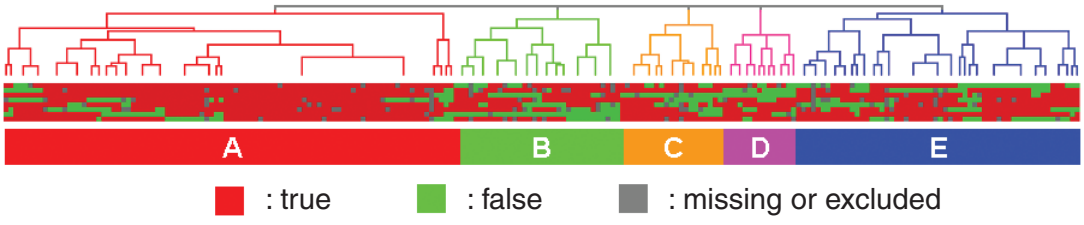


Fig. 4. Kaplan-Meier plot of 4-year overall survival based on the International Prognostic Index and the FNN grouping. Panel A shows a Kaplan-Meier plot of low-clinical-risk patients (IPI score 0–1) grouped on the basis of the FNN. Panel B shows a Kaplan-Meier plot of intermediate-clinical-risk patients (IPI score 2–3) grouped on the basis of the FNN. Panel C shows a Kaplan-Meier plot of high-clinical-risk patients (IPI score 4–5) grouped on the basis of the FNN. The tick marks represent censored data.

A



B

		A (N=92)	B (N=34)	C (N=21)	D (N=15)	E (N=58)	total (N=220)	P Value
FNN model with the highest accuracy		No. 9	No. 6	No. 3	No. 7	No. 2	No. 7	
Accuracy by FNN model shown above		97%	97%	100%	93%	97%	78%	
survival rate	2-year	83%	76%	75%	86%	72%	79%	0.987
	4-year	58%	62%	38%	47%	50%	54%	0.752
IPI	Low	28%	45%	50%	64%	30%	35%	0.157
	Intermediate	54%	42%	44%	29%	55%	50%	
	High	18%	13%	6%	7%	15%	15%	
cell origin	GCB	45%	59%	48%	53%	45%	48%	0.147
	Type3	18%	12%	43%	20%	22%	21%	
	ABC	37%	29%	10%	27%	33%	31%	

Fig. 5. Subgroups of DLBCL by hierarchical clustering with the predictions of FNN models. Panel A shows 220 patients grouped by hierarchical clustering. The dendrogram indicates relatedness of gene expression in each sample, and is color-coded according to sample category. Each column represents a patient, and each row represents a prediction by an FNN model. Red indicates that the prediction is true; green indicates false, and gray indicates missing or excluded. Panel B shows the optimal model, accuracy of the model, and patients' characteristics.

by IPI in those groups. There were no significant characteristics that could be used to designate cell-of-organ groups. The patients predicted correctly by the Cox proportional-hazards model were compared with those predicted correctly by the FNN models. We found that FNN model 6 predicted patient outcome of group B with high accuracy, whereas the Cox model did not (Fig. 6). The Cox model was constructed by Rosenwald *et al*. using 4 signatures: germinal-center B-cell, MHC class II, lymph-node and proliferation.[16] FNN model 6 did not include these 4 signatures, whereas the other models included genes belonging the same clusters as all 4 signatures mentioned above.

Relationships among the selected genes and clinical outcome. Using the FNN models, relationships between input genes and clinical outcome were investigated in each group. One of the attractive features of FNN models is that causal relationships can be described explicitly as IF-THEN rules.[13–15] Firstly, in the constructed FNN model, the connection weights have been

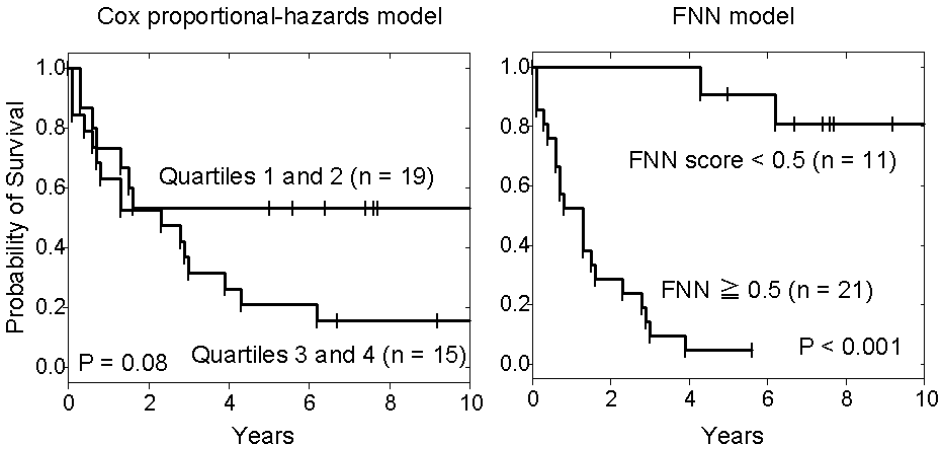


Fig. 6. Kaplan-Meier plot of patients in group B based on Cox proportional-hazards model and the FNN model. The overall survival for patients grouped by the Cox proportional-hazards model[16] (left) and by the FNN model (right) is shown. The tick marks represent censored data.

cluster A				Unknown A			
				L		H	
				nuclear receptor subfamily 3			
				L	H	L	H
MAX dimerization	L	Unknown B	L	7, 34, 51, 57, 76, 88, 98, 100, 109, 124, 130, 177, 238, 281 286, 320, 324	**199**	3, 24, 63, 68, 95, 97, 227, 322, **48, 66, 117, 230**	
			H	**62, 141, 419**	86, 399, 432	**263**	
	H		L	**134, 244, 262, 266, 285**	**189**	**30, 64, 103, 166, 191, 203, 212**	**5, 8, 11, 12, 13, 15, 61, 89, 101, 107, 110, 120, 144, 152, 163, 179, 198, 205, 208, 249, 255, 313, 317, 415**
			H	297, 433, **391**	118, 136, 186, 248, 294, 298, 394	**300, 404**	**78, 411**

Fig. 7. Relationship between expression of 4 genes selected for model 9 and clinical outcome in group A. Because the expression level of each gene can be classified as high or low based on fuzzy reasoning, this model comprised 16 ($=2^4$) fuzzy rules. Light gray areas represent predicted poor prognosis by the FNN. Dark gray areas represent the poorest prognosis. The numbers in each matrix are the patient serial numbers previously described by Rosenwald *et al*.[16] Bold-type underlined numbers indicate the patients who died during the 4-year observation period, and light faced numbers indicate those who survived for at least 4 years. Patient numbers were assigned to cells according to expression levels of each patient.

described in a table consisting of cells with combinations of gene expression levels, e.g. high expression of gene A and low expression of gene B. Next, from comparison of the connection weights, the IF-THEN rules can be defined explicitly as a cause-and-effect chain without any biological or functional experiments. For example, in the previous paper,[13] a poor outcome was predicted when *CD10* expression was low and *IRF-4* expression was high. Fourteen of the patients were identified as having a poor prognosis on the basis of these two factors, and this corresponds to 67% of all patients with poor prognosis.

Relationships in group A, which contained half of the 220 patients, are shown in Fig. 7. From this matrix, simple and precise rules were obtained. Patients who satisfied any 2 rules (e.g., high expression of MAX dimerization with high expression of unknown A, or high expression of MAX dimerization with low expression of unknown B) were predicted to have a poor outcome in the FNN model. One-third of all non-surviving patients (41/118) and 80.4% (41/51) of non-surviving patients in group A satisfied any 2 rules. Detailed analysis of survival time revealed a rule concerning short survival time. High expression of unknown A, nuclear receptor subfamily 3 and MAX dimerization, with low expression of unknown B, was observed in 24 patients. This is a result of intersection between the above 2 rules. Of these 24 patients, 23 died within 2 years, and the remaining patient died at 2 years and 4 months. Therefore, patients who satisfy this rule have significantly high clinical risk; moreover, they included patients with a low IPI score. This shows that these rules (which were defined in terms of gene expression) were independent of IPI.

Next, we investigated gene expression relationships in group E, the second largest group (Fig. 8). High expression of protein tyrosine phosphatase, non-receptor type 2, was observed in 23 patients. Among them, the 18 patients with high expression of oxysterol binding protein-like 3 died during the 4-year observation period. This gene expression rule identifies a tendency toward poor outcome. The activated B-like DLBCL signature includes high expression of protein tyrosine phoshatase, non-receptor type 2, and has poor prognosis.[11, 16] The rule described above, obtained from the FNN model, agrees with this result from prior knowledge. Also, the FNN model identified a rule for favorable outcome, i.e., if expression of oxysterol binding protein-like 3 is low and expression of CD81 is low, the patient will survive. This rule was applied to cases that could not be predicted using well-known prior knowledge.

Discussion

We have developed FNN modeling for prognostic prediction of the heterogeneous cancer DLBCL. Using only 35 genes, a high discrimination ratio (95.9%) and a high prediction accuracy (90.5%) were achieved (Fig. 1). The 4-year survival rate was 91% for patients predicted to survive by the FNN model, and only 10% for those predicted not to survive (Fig. 3). Convenient, high throughput diagnosis may become available in the future if assay systems such as multiplexed quantitative reverse-transcriptase polymerase chain reaction or customized DNA microarrays are established based on expression of the 35 genes we selected. The present method, which uses the expression ratios of these 35 genes and the 9 prediction models, can accurately identify DLBCL patients who are unlikely to be cured by conventional therapy.

The 35 genes selected by the SWEEP operator method included genes that are informative in prognosis of DLBCL. ESTs (*AA805575*) was selected as the first-ranked gene by the SWEEP operator method (Table 1), and the expression pattern of this gene coincided with those of *CD10* and *BCL-6* (Fig. 2). ESTs is the germinal-center B-cell signature, and is associated with favorable outcome.[16] Furthermore, 7 of the 9 gene combinations we used include genes that correlate with known prognostic factors (Fig. 2). These results indicate that the FNN method can select single genes that significantly affect prognosis. However, the FNN method can do more than simply select single genes. It can also construct gene combinations that are likely to produce accurate predictions. Actually, the FNN model with the first-ranked gene selected combinations containing germinal-center B-cell marker and lymph-node marker and proliferation signatures, without prior knowledge.[11, 16] Moreover, by analyzing the connection weight of the FNN model, we can describe causal relationships between gene expression and patient outcome. This ability to select genes and to describe causal relationships may help elucidate DLBCL tumorigenesis.

The notable fact about the present selected combinations of genes is that there is a combination not including genes correlated with the known prognostic factors. This indicates that there are a considerable numbers of prognosis-related genes in addition to the known factors, and that clinical risk of DLBCL patients cannot be predicted using only the known factors.

In our previous study of outcome prediction for 40 DLBCL patients,[13] the gene combination of *CD10* and *IRF-4* was useful

cluster E		① protein tyrosine phosphatase, non-receptor type 2			
		L		H	
		② CD81			
		L	H	L	H
③oxysterol binding protein-like 3	L	1, 14, 94, 116, 173, 247	**159, 184, 216, 223, 423**	72, 114, 279, 401, 430	
	H	**44, 50, 79, 112, 132, 409**	4, 16, 17, 119, 127, 129, 146, 160, 182, 252, 299, 309, 315, 321, 397, 406, 418, 425	**22, 38, 40, 58**	**18, 93, 178, 185, 222, 245, 275, 304, 400, 413, 427, 429, 431, 439**

Fig. 8. Relationship between expression of 3 genes selected for model 2 and clinical outcome in group E. The matrices were constructed in the same way as in Fig. 7. Light gray areas correspond to rules from prior knowledge. Dark gray areas represent new findings.

for outcome prediction. The expression pattern of *IRF-4* correlated with that of *BCL-2*. In the present study, FNN model 7 contained the *CD10*-like gene and *IRF-4*-like gene. It should be noted that model 7 achieved the highest overall accuracy: 78% (Fig. 5). In the other noninferior prediction models constructed in the present paper, about 70% prediction accuracy has been obtained. It is also important that about 22% of patients who were not predicted correctly by model 7 were predicted correctly by other FNN models, which consisted of different genes (Fig. 5). This suggests that DLBCL patients can be divided into subtypes in which prognostic factors distinctly differ from each other and causal relationships also differ, based on the prediction results of FNN models.

In the present study, we classified DLBCL patients based on prediction accuracy of FNN models. The patients in group B could not be predicted successfully with the known gene-expression signatures. However, model 6, which included genes other than the known signatures, produced accurate predictions for group B (Fig. 5B). As shown in Fig. 5, outcome can be accurately predicted for the 5 subgroups using FNN models 9 (A), 6 (B), 3 (C), 7 (D) and 2 (E). The FNN methods can also describe the dominant rule of gene expression for each subtype. Classification without prior knowledge and description of rules may help to elucidate clinical and molecular heterogeneity.

In conclusion, we have developed a method using multiple noninferior FNN models, and have achieved highly accurate prediction outcome of the heterogeneous disease DLBCL. The present paper is the first one describing the multiple noninferior FNN modeling system. The accuracy of outcome prediction did not become lower in spite of using a large number of patients. This approach could be useful for personalized medicine for patients with DLBCL. However, in order to apply the 35 genes to clinical diagnosis, it is necessary to verify the accuracy of the outcome prediction using larger numbers of test cases. In addition, we successfully classified patients based on prediction results. In the future, accurate prognosis and classification of patients using advanced FNN methods may help to individualize a heterogeneous disease and to establish molecularly targeted therapies.

1. A clinical evaluation of the International Lymphoma Study Group classification of non-Hodgkin's lymphoma. The Non-Hodgkin's Lymphoma Classification Project. *Blood* 1997; **89**: 3909–18.
2. The International Non-Hodgkin's Lymphoma Prognostic Factors Project. A predictive model for aggressive non-Hodgkin's lymphoma. *N Engl J Med* 1993; **329**: 987–94.
3. Fisher RI, Gaynor ER, Dahlberg S, Oken MM, Grogan TM, Mize EM, Glick JH, Coltman CA Jr, Miller TP. Comparison of a standard regimen (CHOP) with three intensive chemotherapy regimens for advanced non-Hodgkin's lymphoma. *N Engl J Med* 1993; **328**: 1002–6.
4. Cerny T, Betticher D. Role of high-dose therapy in diffuse large B-cell lymphoma. *Ann Oncol* 2000; **11** Suppl 3: 117–21.
5. Popat U, Przepiork D, Champlin R, Pugh W, Amin K, Mehra R, Rodriguez J, Giralt S, Romaguera J, Rodriguez A, Preti A, Andersson B, Khouri I, Claxton D, De Lima M, Donato M, Anderlini P, Gajewski J, Cabanillas F, van Besien K. High-dose chemotherapy for relapsed and refractory diffuse large B-cell lymphoma: mediastinal localization predicts for a favorable outcome. *J Clin Oncol* 1998; **16**: 63–9.
6. Gatter CK, Warnke AR. Diffuse large B-cell lymphoma. In: Jaffe ES, Harris NL, Stein H, Vardiman JW, editors. World Health Organization classification of tumors. Pathology and genetics of tumors of haematopoietic and lymphoid tissues. Lyon: IARC Press; 2001. p. 157–60.
7. Armitage JO, Weisenburger DD. New approach to classifying non-Hodgkin's lymphomas: clinical features of the major histologic subtypes. Non-Hodgkin's Lymphoma Classification Project. *J Clin Oncol* 1998; **16**: 2780–95.
8. Engelhard M, Brittinger G, Huhn D, Gerhartz HH, Meusers P, Siegert W, Thiel E, Wilmanns W, Aydemir U, Bierwolf S, Griesser H, Tiemann M, Lennert K. Subclassification of diffuse large B-cell lymphomas according to the Kiel classification: distinction of centroblastic and immunoblastic lymphomas is a significant prognostic risk factor. *Blood* 1997; **89**: 2291–7.
9. Harada S, Suzuki R, Uehira K, Yatabe Y, Kagami Y, Ogura M, Suzuki H, Oyama A, Kodera Y, Ueda R, Morishima Y, Nakamura S, Seto M. Molecular and immunological dissection of diffuse large B cell lymphoma: CD5+, and CD5− with CD10+ groups may constitute clinically relevant subtypes. *Leukemia* 1999*;* **13**: 1441–7.
10. Yamaguchi M, Seto M, Okamoto M, Ichinohasama R, Nakamura N, Yoshino T, Suzumiya J, Murase T, Miura I, Akasaka T, Tamaru J, Suzuki R, Kagami Y, Hirano M, Morishima Y, Ueda R, Shiku H, Nakamura S. *De novo* CD5+ diffuse large B-cell lymphoma: a clinicopathologic study of 109 patients. *Blood* 2002; **99**: 815–21.
11. Alizadeh AA, Eisen MB, Davis RE, Davis RE, Ma C, Lossos IS, Rosenwald A, Boldrick JC, Sabet H, Tran T, Yu X, Powell JI, Yang L, Marti GE, Moore T, Hudson J Jr, Lu L, Lewis DB, Tibshirani R, Sherlock G, Chan WC, Greiner TC, Weisenburger DD, Armitage JO, Warnke R, Levy R, Wilson W, Grever MR, Byrd JC, Botstein D, Brown PO, Staudt LM. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* 2000; **403**: 503–11.
12. Shipp MA, Ross KN, Tamayo P, Weng AP, Kutok JL, Aguiar RC, Gaasenbeek M, Angelo M, Reich M, Pinkus GS, Ray TS, Koval MA, Last KW, Norton A, Lister TA, Mesirov J, Neuberg DS, Lander ES, Aster JC, Golub TR. Diffuse large B-cell lymphoma outcome prediction by gene-expression profiling and supervised machine learning. *Nat Med* 2002; **8**: 68–74.
13. Ando T, Suguro M, Hanai T, Kobayashi T, Honda H, Seto M. Fuzzy neural network applied to gene expression profiling for predicting the prognosis of diffuse large B-cell lymphoma. *Jpn J Cancer Res* 2002; **93**: 1207–12.
14. Ando T, Suguro M, Kobayashi T, Seto M, Honda H. Selection of causal gene sets for lymphoma prognostication from expression profiling and construction of prognostic fuzzy neural network models. *J Biosci Bioeng* 2003; **96**: 161–167.
15. Noguchi H, Hanai T, Honda H, Harrison LC, Kobayashi T. Fuzzy neural network-based prediction of the motif for MHC class II binding peptides. *J Biosci Bioeng* 2001; **92**: 227–31.
16. Rosenwald A, Wright G, Chan WC, Connors JM, Campo E, Fisher RI, Gascoyne RD, Muller-Hermelink HK, Smeland EB, Giltnane JM, Hurt EM, Zhao H, Averett L, Yang L, Wilson WH, Jaffe ES, Simon R, Klausner RD, Powell J, Duffey PL, Longo DL, Greiner TC, Weisenburger DD, Sanger WG, Dave BJ, Lynch JC, Vose J, Armitage JO, Montserrat E, Lopez-Guillermo A, Grogan TM, Miller TP, LeBlanc M, Ott G, Kvaloy S, Delabie J, Holte H, Krajci P, Stokke T, Staudt LM. The use of molecular profiling to predict survival after chemotherapy for diffuse large-B-cell lymphoma. *N Engl J Med* 2002; **346**: 1937–47.
17. Horikawa S, Furuhashi T, Uchikawa Y, Tagawa T. A study on fuzzy modeling using fuzzy neural networks. *Proceedings of International Fuzzy Engineering Symposium '91*, 562–73 (1991).
18. Goodnight J. A tutorial on the SWEEP operator. *Am Stat* 1979; **33**: 149–58.
19. Khan J, Wei JS, Ringner M, Saal LH, Ladanyi M, Westermann F, Berthold F, Schwab M, Antonescu CR, Peterson C, Meltzer PS. Classification and diagnostic prediction of cancers using gene expression profiling and artificial neural networks. *Nat Med* 2001; **7**: 673–9.
20. Eisen MB, Spellman PT, Brown PO, Botstein D. Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci USA* 1998; **95**: 14863–8.
21. Yau JC, Dabbagh LK, Formenti KS, Coupland RW, Burns BF, Shaw AR. Expression of transmembrane 4 superfamily member, CD9, is related to improved progression-free survival in patients with diffuse non-Hodgkin's lymphoma. *Oncol Rep* 1998; **5**: 1507–11.
22. Barrans SL, Carter I, Owen RG, Davies FE, Patmore RD, Haynes AP, Morgan GJ, Jack AS. Germinal center phenotype and bcl-2 expression combined with the International Prognostic Index improves patient risk stratification in diffuse large B-cell lymphoma. *Blood* 2001; **99**: 1136–43.
23. Lossos IS, Jones CD, Warnke R, Natkunam Y, Kaizer H, Zehnder JL, Tibshirani R, Levy R. Expression of a single gene, BCL-6, strongly predicts survival in patients with diffuse large B-cell lymphoma. *Blood* 2001; **98**: 945–51.