Jpn. J. Cancer Res. **93**, 1207–1212, November 2002

Fuzzy Neural Network Applied to Gene Expression Profiling for Predicting the Prognosis of Diffuse Large B-cell Lymphoma

Tatsuya Ando,[1] Miyuki Suguro,[2] Taizo Hanai,[1, 3] Takeshi Kobayashi,[1] Hiroyuki Honda[1, 4] and Masao Seto[2]

[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

Diffuse large B-cell lymphoma (DLBCL) is the largest category of aggressive lymphomas. Less than 50% of patients can be cured by combination chemotherapy. Microarray technologies have recently shown that the response to chemotherapy reflects the molecular heterogeneity in DLBCL. On the basis of published microarray data, we attempted to develop a long-overdue method for the precise and simple prediction of survival of DLBCL patients. We developed a fuzzy neural network (FNN) model to analyze gene expression profiling data for DLBCL. From data on 5857 genes, this model identified four genes (*CD10, AA807551, AA805611* and *IRF-4*) that could be used to predict prognosis with 93% accuracy. FNNs are powerful tools for extracting significant biological markers affecting prognosis, and are applicable to various kinds of expression profiling data for any malignancy.

Key words: Fuzzy neural network — Microarray — Gene expression profiling — Diffuse large B-cell lymphoma — Prognosis

Diffuse large B-cell lymphoma (DLBCL) accounts for 30–40% of non-Hodgkin's lymphomas and is well known to include clinically and morphologically heterogeneous groups.[1–5] Current chemotherapeutic regimens can cure some patients with DLBCL, but more than half succumb to the disease.[6, 7] Therefore, identification of high-risk DLBCL groups is crucial and long overdue. Clinical prognostic models such as the International Prognostic Index (IPI; age, performance status, stage, number of extranodal sites and serum lactate dehydrogenase) have been used thus far to establish prognosis for DLBCL patients.[8]

The recent development of microarray analysis provides a new tool for establishing the prognoses of various diseases.[9, 10] With microarray data, comparison of gene expression in tissues from distinct patient groups and their normal counterparts may lead to a more comprehensive and detailed understanding of the molecular mechanisms of the disease than can be obtained from general parameters such as clinical stage and age of the patient.[11, 12] Using microarray analysis, Alizadeh *et al.* concluded that DLBCL can be divided into two groups, germinal center B-like DLBCL and activated B-like DLBCL.[11] The latter has been shown to have a poorer outcome than the former. The predictive value of their model was, however, lower with another DLBCL group.[12] This is partly because Alizadeh's analysis was designed for identifying new DLBCL subtypes, rather than for prognosis. Recently, Shipp *et al.* published a DNA microarray analysis of DLBCL patients who had received standard chemotherapy, and they created a prediction model based on a supervised classification algorithm. This model identified 13 out of the 6817 genes analyzed as being significant for prognosis.[12] However, it would be very beneficial to minimize the number of prognostically significant genes.

Artificial neural network (ANN) analysis is a powerful tool for accurately detecting causal relationships.[13] The fuzzy neural network (FNN) is one of the more advanced ANN models, and its most attractive feature is that the causality between input and output variables can be described extremely accurately as linguistic IF-THEN rules from the acquired model. In a previous study, the FNN model was applied to predicting which peptides would bind to the MHC class II molecule to stimulate an immune response, and these peptides were identified from non-binding peptides with an accuracy of more than 90%.[14] The physicochemical characteristics of peptides with high binding coefficients were described in terms of IF-THEN rules.

In this study, we have applied the FNN method to the separation of clinical DLBCL patients into two groups, those predicted to survive and those predicted to die. We used the gene expression profiling data published by Alizadeh *et al.*[11] From the expression ratios of only four genes, we succeeded in predicting outcomes for 40 DLBCL patients with 93% accuracy.

[3] Present address: Laboratory for Applied Biological Regulation Technology, Graduate School of Bioresource and Bioenvironmental Science, Kyushu University, 6-10-1 Hakozaki, Higashi-ku, Fukuoka 812-8581.

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

MATERIALS AND METHODS

Data preprocessing Expression data of 5857 genes in 40 DLBCL patients were taken from the Stanford Microarray Database (http://genome-www5.stanford.edu/MicroArray/SMD/).[15] Expression data with a fluorescence intensity 1.4 times greater than that of the background were used without alteration, and those with an intensity level less than 1.4 times the background were treated as being at the background level.[11] The ratio of expression intensity of each gene over the reference[11] was used for analysis after having been normalized from 0.1 to 0.9.

FNN modeling A type-I FNN[16] was used to establish the relationship between gene expression and clinical outcome. An FNN combines fuzzy reasoning with an artificial neural network. Construction of an FNN with one input unit and one output unit is described in Fig. 1. Gene expression data, that is, normalized ratio of expression intensity is entered into the FNN and then the data are assigned to two components, high and low, which describe the grade of the gene's expression level. The 'high' and 'low' membership grades are determined by fuzzy inference. The relationship between the 'high' and 'low' membership grades and the output is obtained by training the model using training data, as with ANN.[13, 14, 16, 17] In an FNN model with one input and one output unit, the number of parameters is 6 (two W_c, two W_g and two W_f). In the case of two inputs, it is 12, and $(4n+2^n)$ for an FNN model with n input units and one output unit. The parameters are assigned by model training. To construct an FNN model with many input units, a large set of training data are needed. In the present study, the candidate input variables are the 5857 gene expression data points for each patient. To make the FNN model as simple as possible, the input variables were selected by the SWEEP operator method.[17, 18] In our previous papers,[13, 14, 17] we directly selected the input variables by the parameter increasing method (PIM)[14] among input candidates and simultaneously constructed the FNN model by training. In the present paper, however, the computational time for this method would have been prohibitive. To select the input genes in a short time, the selection was carried out by a SWEEP operator method without training. In this method, briefly, the weight parameters W_c and W_g were fixed and W_f was determined by the SWEEP operator method. Since no training was carried out, the method was suitable for the selection of input variables from the large number of input candidates (5857 genes). As a first step, an FNN model with one input unit was created. Expression data for the gene from all 40 patients' data sets were entered into the FNN model, and the W_f was determined by the SWEEP operator method. We repeated this procedure 5857

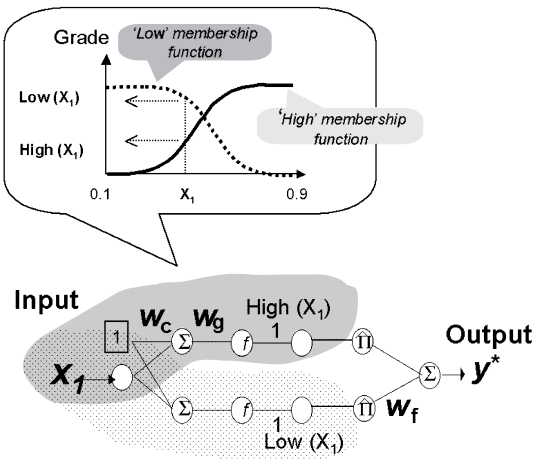

Fig. 1. Structure of the fuzzy neural network (FNN). Gene expression data, in the form of normalized expression ratios, were entered into the FNN and then the data were assigned to two components, 'high' expression (gray curve) and 'low' expression (dotted curve). The 'high' and 'low' membership grades were determined by fuzzy inference based on the actual expression data for each gene (upper panel); when the normalized expression ratio of a gene in a given patient was 0.5, a grade of 0.5 was assigned for both 'high' and 'low' membership as the value before training.

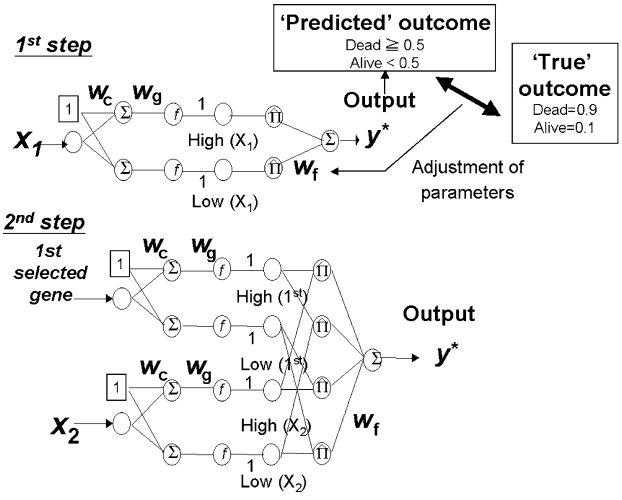

Fig. 2. FNN modeling. As the first step, an FNN model with one input unit was constructed. Expression data for one gene in all 40 patients was entered into the FNN model, and the weight parameter W_f were determined by the SWEEP operator method. This procedure was repeated 5857 times to construct a model for each gene. Then the models were compared for the accuracy with which they predicted patient survival or death, and the gene used in the most accurate model was selected as the "1st gene." Next, the PIM[14] was applied. With the "1st gene" fixed, a similar method was used to select the "2nd gene," which yielded the highest accuracy model in combination with the "1st gene."

times to construct a model for each gene. Then the accuracy of the models was compared, and the gene used in the model with the highest prediction accuracy was selected as the "1st gene." Next, the PIM method was applied. Having the "1st gene" fixed, we used a similar method to select the "2nd gene," which gave the highest accuracy in combination with the "1st gene" (Fig. 2). Having the "1st gene" and the "2nd gene" fixed, we then selected the 3rd gene. We repeated this procedure, increasing the number of input genes one by one. Then, we used training to construct FNN models with the selected genes (Fig. 3).

In the present paper, the threshold value of 0.5 was used for determining the predicted outcome; when the output was more than 0.5, it was considered that the model predicted the death of the patient (Fig. 2). The predicted outcomes were compared with the true outcome, alive or dead, and the training of the model, i.e., the adjustment of the parameters, was done by a back-propagation algorithm. The training ratio and training time were set at 0.1 and 5000 iterations, respectively. The performance of the FNN models were assessed by cross-validation (Fig. 4).[19] In each model, 30 data sets were used for training and 10 for evaluation. Since the number of data sets was not large, the performance of the model depended on which data were used for training. In order to assess the model fairly, such cross-validation was carried out for each model.

Survival analysis and Kaplan-Meier plots The Kaplan-Meier survival analysis plots were computed using StatView for Windows version 5.01 (SAS Institute, Inc., Cary, NC). The differences in survival rates were analyzed by a log-rank test (Mantel-Cox method).

RESULTS AND DISCUSSION

Genes selected for prognosis An FNN model with four genes as input variables was constructed from data on 5857 genes. These four input genes are listed in Table I in the order in which they were selected. The earlier the input variables were selected, the more important they were as input for the predictions. The model could predict the 4-year survival rate for DLBCL patients with an accuracy of 93%. This value is high compared to that from the hierarchical clustering method, 72%.[11]

CD10, a membrane metallo-endopeptidase, was the first variable selected in our analysis. It was more strongly expressed in DLBCL survivors than in those who died. Its expression has also been reported to be associated with a better response to treatment of acute lymphoblastic leukemia (ALL) and neuroblastoma.[20, 21] CD10 is expressed in normal precursor B-cells and follicle center cells.[22, 23] Leukemias and lymphomas originating from those cells, such as ALL and follicular lymphoma, also express CD10. Interestingly, about one-third of DLBCLs are known to express CD10, suggesting that their normal counterpart is the germinal center B-cell.[24] Alizadeh *et al.* reported that the germinal center B-like DLBCLs showed a better outcome than the activated B-like DLBCLs.[11] We next selected two as-yet unidentified genes, and then *IRF-4* (*MUM1*/*LSIRF*). *IRF-4* was identified as a protooncogene

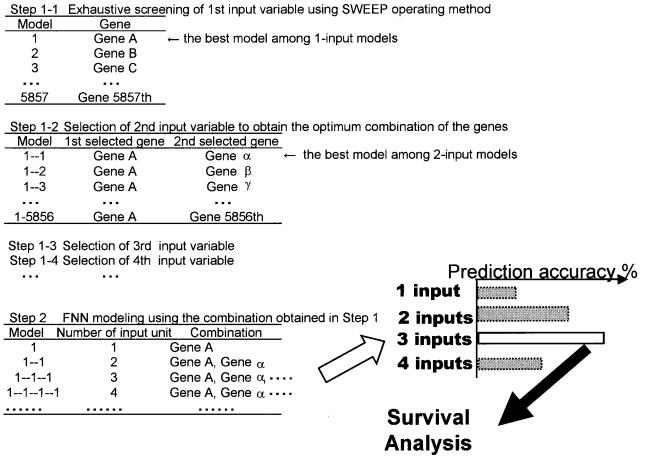

Fig. 3. Strategy of model selection. The combination of 1st and 2nd genes was selected by the SWEEP operator method and the PIM. Having the "1st gene" and the "2nd gene" fixed, we selected the 3rd gene. We repeated this procedure, increasing the number of input genes one by one. Then, FNN models with the selected genes were constructed by training, and the performance of the model was tested.

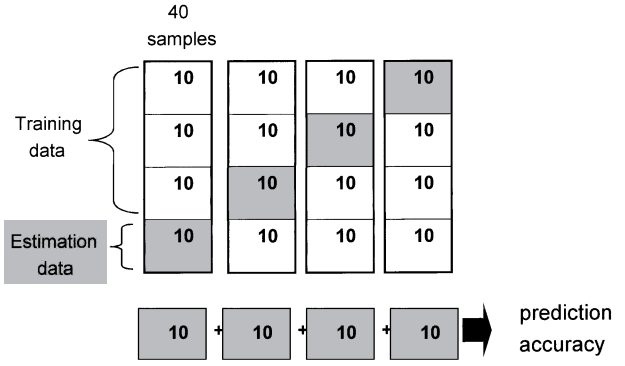

Fig. 4. A cross-validation. A 4-fold cross-validation was carried out to fairly test predictions for 40 patients. The 40 data sets were divided into training data (30 data sets) and evaluation data (10 data sets). The FNN model was then optimized using the 30 training data sets and validated with the 10 evaluation data sets. This training procedure was repeated four times so that the data for each patient were assessed once as evaluation data. Then the prediction accuracy across all four trials was calculated and averaged for the overall accuracy of the FNN model.

activated by chromosome translocation t(6;14)(p25;q32), which juxtaposes the immunoglobulin heavy-chain locus to the *IRF-4* gene in multiple myeloma cell lines.[25] The activated B-like DLBCL signature includes high *IRF-4* expression and has a poorer prognosis.[11] These findings suggest that *IRF-4* expression may confer a growth advantage on the lymphoma cells, thus accounting for their aggressive clinical behavior. *AA807551*, a hypothetical gene also listed as accession number AL512731 in GenBank, was more strongly expressed in those who died than in the survivors of DLBCLs. However, the biological significance of the expression of this gene, and of *AA805611*, remains to be explored. For all four genes, differences in expression between the two patient groups showed a *P* value of less than 0.05 (Table I).

Table I. Four Genes Selected with the FNN Model

Order of selection	Selected genes	*P* value	Predictive accuracy
1	*CD10*	0.008	93%
2	Unknown (*AA807551*)	0.002	
3	Unknown (*AA805611*)	0.032	
4	*IRF-4*	0.022	

The genes selected by the PIM are shown in the order selected. The *P* value for each gene, calculated by the Mann-Whitney test, indicates the significance of expression differences between patients with 4-year survival and those without. The predictive accuracy was determined by cross-validation (Fig. 4).

				1. CD10			
				L		H	
				2. Unknown (AA807551)			
				L	H	L	H
3. Unknown (AA 805611)	L	4. IRF- 4	L		33	1, 9,15,37	**18, 41**
			H	**7, 21, 23, (24) 26, 31**	**2, (5) 6, 12, 16, 25, 42, 49**	3, 14, 28, 32, 40	**13, 27, 34, 48**
	H		L			10	8
			H	39	**17, 36**	4, 30	**(11), 20, 29**

Fig. 5. Relationship among four input genes and predicted outcome. H and L respectively refer to high and low expression level of each gene. Since the expression level of each gene can be divided into either high or low groups according to fuzzy reasoning, this model comprised 16 ($=2^4$) fuzzy rules. Light gray areas (□) represent predicted poorer prognosis by the FNN. Dark gray areas (■) represent the poorest prognosis. Numbers in each matrix cell are the patients' numbers previously described by Alizadeh *et al.*[11] Bold type numbers with underline indicate the patients dead within 4 years, and italic type numbers alive. Patient numbers are placed in the matrix according to the expression levels of the four genes in that patient. Patient numbers in circles represent incorrect classification by the FNN.

A previous report showed that mutations of the *p53* gene were associated with a poor prognosis in aggressive B-cell lymphoma.[26] However, owing to the unavailability of p53 expression data from some patients, we could not include the *p53* gene in the analysis.

Relationships among the selected genes From the constructed FNN model, the relationship between the input of four genes and the output of the survival score is described as a fuzzy rule, shown in Fig. 5. From this matrix, the following rules are obtained. Patients with low expression of CD10 were predicted to have a poor prognosis in the FNN model. A poor outcome was predicted particularly when CD10 expression was low and *IRF-4* expression was high. Fourteen of the patients were identified as having poor prognosis on the basis of these two factors, which corresponds to 67% of all patients with poor prognosis. Patients, No. 5, 24, 33 and 39, are four exceptional cases.

Furthermore, the FNN model also identified cases with a poor prognosis despite a high expression ratio of CD10. The correct identification of these cases was obtained by adding the expression information on the other two genes; the outcomes of patients are poor even though CD10 expression is high if *AA807751* is high and *AA805611* is low. Six patients belong to this causality group without exception.

Our study attained a high prediction accuracy of 93% with the FNN model. This means that three out of 40 patients' prognoses were incorrectly predicted. These three

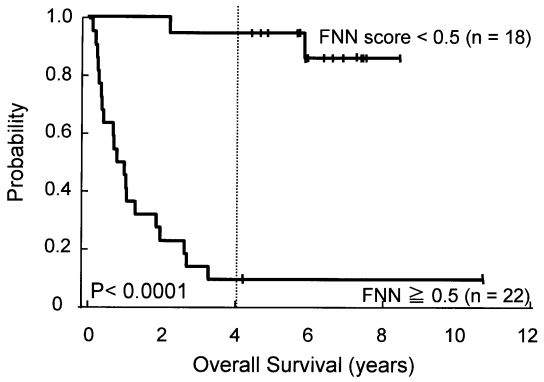

Fig. 6. Kaplan-Meier plot of the 4-year overall survival for all patients grouped by FNN score. The *P* value for the prediction outcome groups is computed using a log-rank test. The tick marks represent censored data.

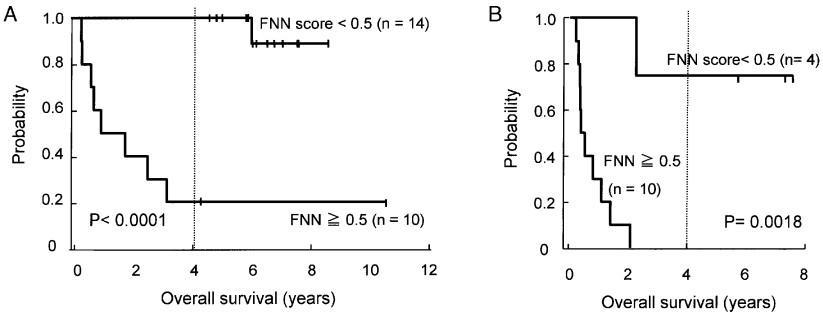

Fig. 7. A. Kaplan-Meier plot of the 4-year overall survival of low clinical risk patients (IPI score 0−2) grouped by FNN score. B. Kaplan-Meier plot of high clinical risk patients (IPI score 3−5) grouped by FNN score. The tick marks represent censored data.

patients, Nos. 5, 11 and 24, are indicated with circles in Fig. 5. Patients No. 11 and 24 had intermediate levels of CD10 expression, indicating the patients were considered marginal for survival.

Kaplan-Meier plots Kaplan-Meier survival analyses indicated that the patients predicted alive by the FNN model showed longer survival than the patients predicted dead (Fig. 6). This result indicates the existence of a gene-expression signature in DLBCL associated with a better outcome. Kaplan-Meier plots of overall survival showed the independence of the IPI and the groups based on the FNN model (Fig. 7). Among the patients whom IPI put in the low-risk group, FNN successfully identified those who had poorer prognosis. Patient prognosis was the poorest for patients in the high-risk group by IPI and predicted dead by FNN (Fig. 7B). Thus, the FNN is more informative in combination with clinical presentation. The patients predicted to have the poorest outcome by both the IPI and the FNN should be treated with a therapy other than conventional chemotherapy.

In conclusion, these results indicate that the FNN model is a powerful tool for identification of genes significant for prognosis. The FNN model should be applicable to microarray analysis data obtained from various malignancies.

ACKNOWLEDGMENTS

This work was supported in part by a Grant-in-Aid for the development of a model system for the organization of biotechnological industry information from the Japan Biological Informatics Consortium, a Grant-in-Aid for the 2nd Term Comprehensive 10-Year Strategy for Cancer Control from the Ministry of Health, Labour and Welfare, and Grants-in-Aid for Science in Primary Areas (Cancer Research) and for the Encouragement of Young Scientists from the Ministry of Education, Culture, Sports, Science and Technology, Japan.

(Received July 22, 2002/Revised September 6, 2002/Accepted September 11, 2002)

REFERENCES

1) Gatter, K. C. and Warnke, R. A. Diffuse large B-cell lymphoma. *In* "Pathology and Genetics of Tumors of Haematopoietic and Lymphoid Tisseus," ed. E. S. Jaffe, N. L. Harris, H. Stein and J. M. Vardimam, pp. 171−174 (2001). IARC Press, Lyon, France.
2) Engelhard, M., Brittinger, G., Huhn, D., Gerhartz, H. H., Meusers, P., Siegert, W., Thiel, E., Wilmanns, W., Aydemir, U., Bierwolf, S., Griesser, H., Tiemann, M. and 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*, **89**, 2291−2297 (1997).
3) Wilson, W. H., Teruya-Feldstein, J., Fest, T., Harris, C., Steinberg, S. M., Jaffe, E. S. and Raffeld, M. Relationship of p53, bcl-2, and tumor proliferation to clinical drug resistance in non-Hodgkin's lymphomas. *Blood*, **89**, 601−609 (1997).
4) Armitage, J. O. and Weisenburger, D. D. New approach to classifying non-Hodgkin's lymphomas: clinical features of the major histologic subtypes. Non-Hodgkin's Lymphoma Classification Project. *J. Clin. Oncol.*, **16**, 2780−2795 (1998).
5) 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. and Nakamura, S. *De novo* CD5+ diffuse large B-cell lymphoma: a clinicopathologic study of 109 patients. *Blood*, **99**, 815−821 (2002).
6) 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. and 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.*, **16**, 63–69 (1998).
7) Cerny, T. and Betticher, D. Role of high-dose therapy in diffuse large B-cell lymphoma. *Ann. Oncol.*, **11** (Suppl. 3), 117–121 (2000).
8) The International Non-Hodgkin's Lymphoma Prognostic Factors Project. A predictive model for aggressive non-Hodgkin's lymphoma. *N. Engl. J. Med.*, **329**, 987–994 (1993).
9) Pomeroy, S. L., Tamayo, P., Gaasenbeek, M., Sturla, L. M., Angelo, M., McLaughlin, M. E., Kim, J. Y., Goumnerova, L. C., Black, P. M., Lau, C., Allen, J. C., Zagzag, D., Olson, J. M., Curran, T., Wetmore, C., Biegel, J. A., Poggio, T., Mukherjee, S., Rifkin, R., Califano, A., Stolovitzky, G., Louis, D. N., Mesirov, J. P., Lander, E. S. and Golub, T. R. Prediction of central nervous system embryonal tumour outcome based on gene expression. *Nature*, **415**, 436–442 (2002).
10) Van't Veer, L. J., Dai, H., Van De Vijver, M. J., He, Y. D., Hart, A. A., Mao, M., Peterse, H. L., Van Der Kooy, K., Marton, M. J., Witteveen, A. T., Schreiber, G. J., Kerkhoven, R. M., Roberts, C., Linsley, P. S., Bernards, R. and Friend, S. H. Gene expression profiling predicts clinical outcome of breast cancer. *Nature*, **415**, 530–536 (2002).
11) Alizadeh, A. A., Eisen, M. B., Davis, R. E., Ma, C., Lossos, I. S., Rosenwald, A., Boldrick, J. C., Sabet, H., Tran, T., Yu, X., Powell, J. I., Yang, L., Marti, G. E., Moore, T., Hudson, J., Jr., Lu, L., Lewis, D. B., Tibshirani, R., Sherlock, G., Chan, W. C., Greiner, T. C., Weisenburger, D. D., Armitage, J. O., Warnke, R., Levy, R., Wilson, W., Greverss, M. R., Byrd, J. C., Bostein, D., Brown, O. P. and Staudt, L. M. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature*, **403**, 503–511 (2000).
12) Shipp, M. A., Ross, K. N., Tamayo, P., Weng, A. P., Kutok, J. L., Aguiar, R. C., Gaasenbeek, M., Angelo, M., Reich, M., Pinkus, G. S., Ray, T. S., Koval, M. A., Last, K. W., Norton, A., Lister, T. A., Mesirov, J., Neuberg, D. S., Lander, E. S., Aster, J. C. and Golub, T. R. Diffuse large B-cell lymphoma outcome prediction by gene-expression profiling and supervised machine learning, *Nat. Med.*, **8**, 68–74 (2002).
13) Hanai, T., Hibino, S., Nagata, E., Matsubara, M., Fukagawa, K., Shirataki, T., Honda, H. and Kobayashi, T. Assessment of senile dementia of Alzheimer type using artificial neural networks. *Jpn. J. Med. Electron. Biol. Eng.*, **37**, 178–183 (1999) (in Japanese).
14) 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).
15) Sherlock, G., Hernandez-Boussard, T., Kasarskis, A., Binkley, G., Matese, J. C., Dwight, S. S., Kaloper, M., Weng, S., Jin, H., Ball, C. A., Eisen, M. B., Spellman, P. T., Brown, P. O., Botstein, D. and Cherry, J. M. The Stanford Microarray Database. *Nucleic Acids Res.*, **29**, 152–155 (2001).
16) Horikawa, S., Furuhashi, T., Uchikawa, Y. and Tagawa, T. A study on fuzzy modeling using fuzzy neural networks. *Proc. Int. Fuzzy Eng. Symp., '91*, 562–573 (1991).
17) Noguchi, H., Hanai, T., Takahasi, W., Ichii, T., Tanikawa, M., Masuoka, S., Honda, H. and Kobayashi, T. Model construction for quality of beer and brewing process using FNN. *J. Chem. Eng. Jpn.*, **25**, 695–701 (1999).
18) Goodnight, J. A tutorial on the SWEEP operator. *Am. Stat.*, **33**, 149–158 (1979).
19) Khan, J., Wei, J. S., Ringner, M., Saal, L. H., Ladanyi, M., Westermann, F., Berthold, F., Schwab, M., Antonescu, C. R., Peterson, C. and Meltzer, P. S. Classification and diagnostic prediction of cancers using gene expression profiling and artificial neural networks. *Nat. Med.*, **7**, 673–679 (2001).
20) Mandel, M., Rechavi, G., Neumann, Y., Biniaminov, M., Rosenthal, E., Toren, A., Brok-Simoni, F., Ben-Bassat, I. and Ramot, B. CD10+ cell population in the bone marrow of patients with advanced neuroblastoma. *Med. Pediatr. Oncol.*, **22**, 115–118 (1994).
21) Uckun, F. M., Sather, H., Gaynon, P., Arthur, D., Nachman, J., Sensel, M., Steinherz, P., Hutchinson, R., Trigg, M. and Reaman, G. Prognostic significance of the CD10(+) CD19(+) CD34(+) B-progenitor immunophenotype in children with acute lymphoblastic leukaemia: a report from the Children's Cancer Group. *Leuk. Lymphoma*, **27**, 445–457 (1997).
22) Shipp, M. A., Richardson, N. E., Sayre, P. H., Brown, N. R., Masteller, E. L., Clayton, L. K., Ritz, J. and Reinherz, E. L. Molecular cloning of the common acute lymphoblastic leukemia antigen (CALLA) identifies a type II integral membrane protein. *Proc. Natl. Acad. Sci. USA*, **85**, 4819–4823 (1988).
23) Arber, D. A. and Weiss, L. M. CD10: a review. *Appl. Immunohistochem.*, **5**, 125–140 (1997).
24) Harada, S., Suzuki, R., Uehira, K., Yatabe, Y., Kagami, Y., Ogura, M., Suzuki, H., Oyama, A., Kodera, Y., Ueda, R., Morishima, Y., Nakamura, S. and Seto, M. Molecular and immunological dissection of diffuse large B cell lymphoma: CD5+, and CD5− with CD10+ groups may constitute clinically relevant subtypes. *Leukemia*, **13**, 1441–1447 (1999).
25) Iida, S., Rao, P. H., Butler, M., Butler, M., Corradini, P., Boccadoro, M., Klein, B., Chaganti, R. S. and Dalla-Favera, R. Deregulation of MUM1/IRF4 by chromosomal translocation in multiple myeloma. *Nat. Genet.*, **17**, 226–230 (1997).
26) Ichikawa, A., Kinoshita, T., Watanabe, T., Kato, H., Nagai, H., Tsushita, K., Saito, H. and Hotta, T. Mutations of the p53 gene as a prognostic factor in aggressive B-cell lymphoma. *N. Engl. J. Med.*, **337**, 529–534 (1997).