Clinical Proteomics Journal
Copyright © Humana Press Inc.
All rights of any nature whatsoever are reserved.
ISSN 1542-6416/04/01:365–373/$25.00

Original Article

Serum Proteomic Pattern for Predicting Recurrence of Undifferentiated Nasopharyngeal Carcinoma After Radiotherapy

A Potential Surrogate Marker

*Terence C. W. Poon, Shuk Man Chow, Brigette B. Y. Ma, Edwin P. Hui, and Anthony T. C. Chan**

Department of Clinical Oncology, The Chinese University of Hong Kong, The Prince of Wales Hospital, Shatin, N.T., Hong Kong SAR, China

Abstract

Although most patients with early-stage nasopharyngeal carcinoma (NPC) can be cured by radiotherapy, there is a high recurrence rate in patients with advanced NPC. We attempted to identify proteomic patterns in sera for predicting tumor recurrence. Pretreatment sera were collected from 64 NPC patients with complete remission after radiotherapy. Serum proteins were profiled by SELDI ProteinChip technology, and correlated with local/distant recurrence.

Forty proteomic features were significantly different between the patient groups with and without tumor recurrence. Univariate analyses showed that 32 of them were significantly associated with time to first recurrence. Multivariate Cox-regression analyses identified International Union Against Cancer (UICC) stage and two proteomic features with mass/charge (m/z) values of 8808 and 6626 as independent prognostic indicators for tumor recurrence. The hazard ratios were 2.0 (95% confidence interval, CI 1.3–3.2) and 0.79 (95% CI 0.64–0.96) for a double of peak intensity of proteomic feature m/z 8808 and m/z 6626, respectively. These two proteomic features were also independent prognosticators for overall survival. A decision tree was constructed to predict the tumor recurrence by using UICC stage, proteomic feature m/z 8808, and proteomic feature m/z 6626, and evaluated by Leave-One-Out cross-validation. Kaplan-Meier analysis confirmed that the decision tree could predict both recurrence-free survival and overall survival. The positive and negative predictive values for tumor recurrence within 4 yr were 74 and 89%, respectively.

A serum proteomic pattern comprising features m/z 8808 and m/z 6626 is a potential surrogate marker of disease recurrence after radiotherapy in NPC.

*Author to whom all correspondence and reprint requests should be addressed:
Email: anthonytcchan@cuhk.edu.hk

Key Words: SELDI-TOF MS; ProteinChip array; prognosis; decision tree; survival.

Introduction

Nasopharyngeal carcinoma (NPC) is an endemic disease in Southern China *(1)*. Over 60% of NPC patients present with advanced International Union Against Cancer (UICC) stage (stages 3 and 4) disease. Primary treatment for non-metastatic NPC is radical external radiotherapy (RT). Most patients with early-stage NPC will achieve a complete clinical remission after RT with or without concurrent chemotherapy. However, patients with locoregionally advanced disease (UICC stages 3 and 4) have significant recurrence rates, 30–40%, of both local recurrences and particularly distant metastases following RT alone *(2)*. Once the patients have developed local recurrences or distant metastases, the treatment outcome is poor. If such patients with a high risk of recurrence were identified, more intensive therapy could be provided to improve the treatment outcomes, and those with a lower risk of recurrence could be spared from such a potentially toxic treatment.

Surface-enhanced laser desorption/ionization (SELDI) ProteinChip technology is a recent proteomic tool that has been applied to the discovery of diagnostic proteomic patterns for various diseases including cancers and infectious diseases *(3–6)*. Apart from application in cancer diagnosis, recent studies have strongly suggested that the SELDI ProteinChip technology is also useful in discovering proteomic pattern for monitoring recurrence of NPC *(7)* and predicting treatment response in cancer patients *(8)*. In the present study we aimed to identify serum proteomic features by the SELDI ProteinChip technology for predicting local/distant recurrence in NPC patients after curative RT.

Material and Methods

Patients

This study was approved by the Ethics Committee of the Chinese University of Hong Kong. Pretreatment serum samples, that were stored at –70ºC, were archived for this retrospective study. The samples were from 64 patients with newly diagnosed undifferentiated NPC (World Health Organization type III histology) admitted between October 1995 and July 1997 at the Department of Clinical Oncology, Prince of Wales hospital, Hong Kong. Disease staging was based on the UICC 1997 (5th Edition) classification, and all patients underwent pretreatment evaluation with physical examination, chest X-ray, contrast-enhanced computed tomography of the nasopharynx, and nasopharyngoscopy. Patients with UICC tumor stage (T) 3 to 4 and/or node stage (N) 2–3 disease also had baseline bone scan and abdominal ultrasound (US). External RT was administered to all patients according to the techniques described by Ho et al. *(1)*. Standard fractionated RT at 2 gy/d was delivered to the primary tumor and nodes to a total dose of 66–70 gy. Twelve patients also received concurrent cisplatin and RT (CT-RT) as part of a clinical trial, at a dose of 40 mg/m^2 over a weekly schedule *(11)*. All 64 patients achieved complete remission by 4–6 wk after completion of RT, and were followed thereafter every 8–12 wk with physical examination and nasopharyngoscopy until February 3, 2001. Suspected disease recurrence were investigated radiologically with CT, US or magnetic resonance imaging (MRI), and confirmed histologically where possible.

Serum Proteomic Profiling

Serum samples were subjected to the SELDI ProteinChip analysis to obtain a quantitative proteomic profile with molecular mass

ranging from 0.9 to 250 kDa, as previously described *(4,5)*. The intra- and inter-assay coefficient of variations of the peak intensities were less than 15%. All samples were analyzed blindly without knowledge of the clinical information. It was performed with two types of ProteinChip Arrays, H50 and CM10. For the hydrophobic H50 ProteinChip assay, the H50 ProteinChip arrays were pre-equilibrated with 150 μL of 0.1% trifluoroacetic acid (TFA) solution, and subsequently with 150 μL of sample diluent (phosphate-buffered saline containing 0.1% TFA) for 5 min. Five microliters of 20-fold diluted serum was added to the ProteinChip arrays in duplicate. For the weak cation exchange CM10 ProteinChip assay, 2 μL of the serum sample were denatured by adding 4 μL of U9 solution (9 *M* urea, 2% CHAPS, 50 m*M* Tris-HCl, pH 9.0), and subsequently diluted with 34 μL of T4 solution (50 m*M* sodium acetate, 0.1% Trition X-100, pH 4.0) to result in a final dilution of 20-fold. CM10 ProteinChip arrays were pre-equilibrated twice with 150 μL of the binding buffer for 5 min. Then 5 μL of the diluted sample were applied to the ProteinChip array in duplicate. For both H50 and CM10 ProteinChip assays, serum proteins were allowed to bind to the arrays with shaking at room temperature for 90 min. After the incubation, the H50 and CM10 ProteinChip arrays were washed five times with 0.1% TFA solution and T4 solution, respectively. After rinsing with deionized water and air-drying, sinapinic acid matrix in 50% acetonitrile and 0.5% TFA was added to each array. The ProteinChip Arrays were read on a ProteinChip PBS II reader of the ProteinChip Biomarker System to measure the masses and intensities of the protein peaks (Ciphergen Biosystems). The common peaks among the SELDI mass spectra were identified and quantified using the Biomarker Wizard software (Ciphergen Biosystem). The peak intensities were normalized with the total ion current, and subsequently with the total peak intensities. Before data mining, the normalized peak intensities of the duplicate measurements were averaged, followed by $\log_2$ transformation.

Identification of Recurrence-Associated Proteomic Features by Statistical Analysis

To avoid identification of proteomic features that might be only associated with the investigated patients, but not recurrence itself, recurrence-associated proteomic features were the SELDI proteomic features that fulfilled three criteria: (1) normalized peak intensity was statistically significantly different between the patients with and without tumor recurrence; (2) the proteomic feature was statistically significantly associated with the time to first recurrence as shown by univariate analysis; and (3) the proteomic feature was an independent prognostic indicator as shown by multivariate analysis.

Statistical analyses were performed by SPSS program release 10.0 (SPSS Inc., Chicago, IL). To achieve a higher stringency of the statistical tests, for univariate statistical analyses a null hypothesis was rejected when a two-sided *p*-value was less than 0.025. The difference in a proteomic feature between the patient groups with and without local/distant recurrences was evaluated using the Mann-Whitney U-test. Recurrence-free survival was defined from the date of completion of radiotherapy to the date of first local/distant recurrence or censored at the date of last follow up. Overall survival was defined from the date of confirmed diagnosis to the date of death or last follow-up. Univariate Cox-regression analysis was applied to evaluate the associations of the proteomic features with time to first local/distant recurrence. A stepwise Cox proportional hazard model was used to examine the association of various prognostic factors, including the proteomic features, age, sex, T stage, N stage, and UICC overall stage.

Construction and Evaluation of Decision Tree for Predicting Recurrence

A Classification and Regression Tree algorithm (CART) model as implemented in S-PLUS (Insightful, Seattle, WA) *(9)* was trained to construct a decision tree for predicting NPC patients with or without local/distant recurrences, as previously described by Poon et al. *(10)* and Markey et al. *(11)*. Independent prognostic indicators identified by the multivariate Cox-regression analysis were used as input variables for generating the tree model. To prevent over-fitting, the decision tree was pruned to reduce terminal branches to 5 or fewer. The purpose of building a decision tree is to provide a parsimonious, and clinically easy-to-use, prediction system.

The performance of the resulted decision tree model was evaluated by Leave-One-Out cross-validation *(11,13)*. Briefly, a CART model was trained on (64–1) cases, and the trained model was then used to test the case that had been left out. This process was repeated 64 times until every case in the dataset had been used once as an unseen independent test case. The results from the 64 test cases were used to estimate the classifier's prediction performance. The prognostic significance of the model on recurrence-free survival and overall survival was tested by the univariate Kaplan-Meier method and Log rank test. To assess the performance of the decision tree as a prediction tool for important clinical outcomes, the positive and negative predictive values of the model was calculated using standard formulae.

Results

Overall Clinical Outcome

The patient characteristics of the 64 patients are listed in Table 1. Data collected up to February 3, 2001 was included in the survival analysis. The median follow-up was 3.9 yr. At the time of last follow-up, 14 patients developed local recurrences, and 13 developed distant metastases. Three had both local and distant metastases, so that there was a total of 24 recurrence events. Thirty-five patients were alive without disease; one patient was alive with disease. There were 28 deaths, 23 of which were related to tumor recurrence.

Serum Proteomic Features Associated With Patients With Tumor Recurrence

A total of 690 proteomic features were found among the pretreatment serum samples, 331 by the hydrophobic H50 ProteinChip arrays and 359 by the weak cation ion exchange CM10 ProteinChip arrays. To search for proteins/polypeptides significantly associated with recurrence, we performed Mann-Whitney U-test to compare proteomic features between cases with and without local/distant recurrence. Relative levels of 22 and 18 features were found to be significantly higher and lower in the group with tumor recurrence, respectively (all p-values < 0.025). Univariate Cox-regression analysis showed that 32 of these 40 proteomic features were significantly associated with time to first tumor recurrence (all p-values < 0.025).

Multivariate Analyses of Recurrence-Free Survival and Overall Survival

Multivariate Cox-regression analysis was performed to examine whether the proteomic features could provide prognostic information in additional to the traditional TNM staging system. The multivariate Cox-regression model showed that among the 32 significant proteomic features, proteomic features with mass/charge (m/z) values of 8808 (range 8796–8821) ($p = 0.002$) and 6626 (range 6623–6629) ($p = 0.02$), and overall UICC stage ($p = 0.003$) were independent prognostic indicators of tumor recurrence. Sex, age, T stage, and N stage were not statistically significant, and excluded in the model. The multivariate Cox-regression model of recurrence-free survival revealed hazard ratios of 2.0 (95%

confidence interval, CI 1.3–3.2) for a double of the peak intensity of the proteomic feature m/z 8808, 0.79 (95% CI 0.64–0.96) for a double of the peak intensity of the proteomic feature m/z 6626, and 2.2 (95% CI 1.3–3.7) for a UICC overall stage increment. The multivariate Cox-regression analysis also showed that proteomic feature m/z 8808 ($p = 0.046$), proteomic feature m/z 6626 ($p = 0.013$), and UICC overall stage ($p = 0.001$) were statistically significantly associated with overall survival. Therefore, our data indicated that higher serum level of proteomic feature m/z 8808 and lower serum level of proteomic feature m/z 6626 were significantly associated with higher rate of tumor recurrence and poor overall survival, and were independent predictors in additional to overall UICC stage (Fig. 1).

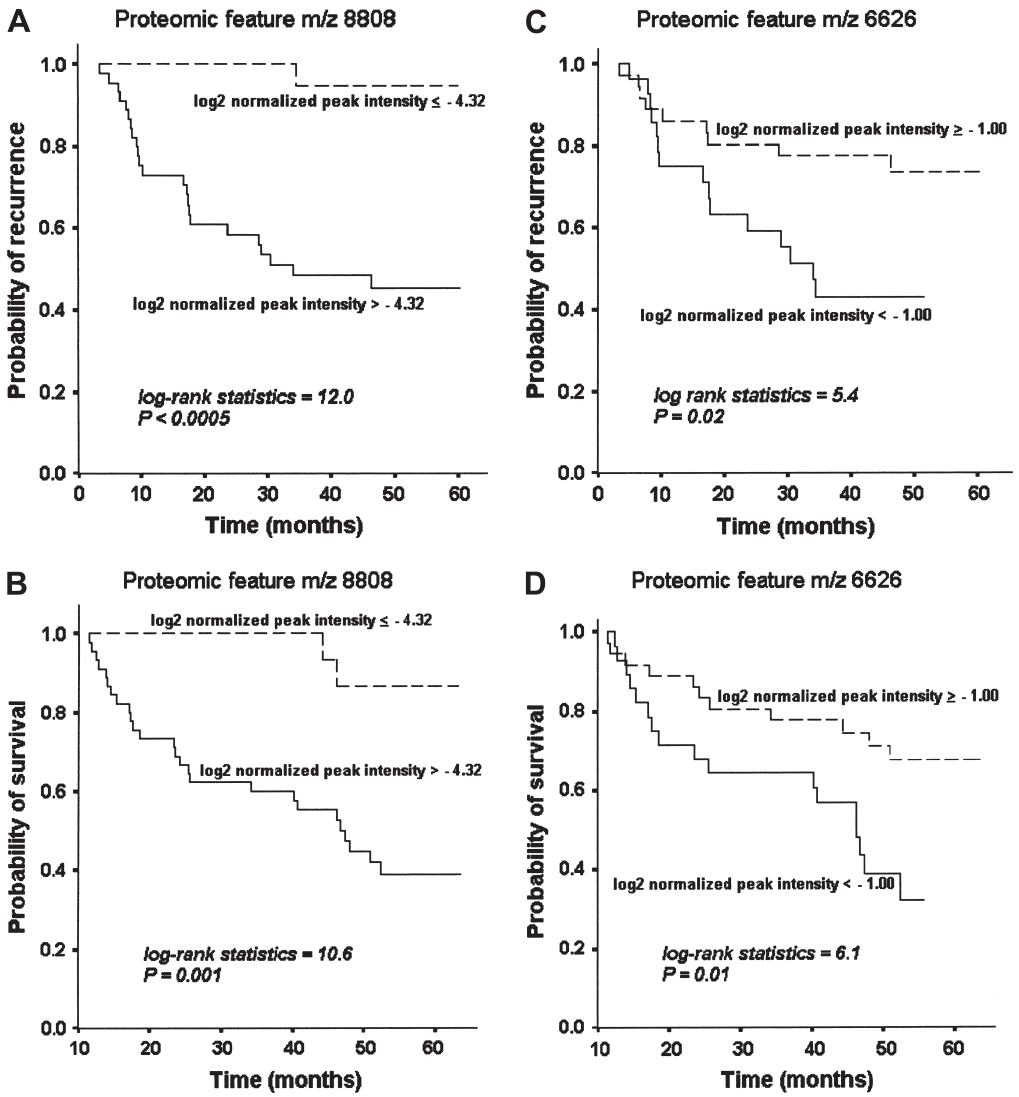

Fig. 1. Higher peak intensity of proteomic feature mass/charge (m/z) 8808 was significantly associated with higher rate of tumor recurrence **(A)** and poor overall survival **(B)**, whereas lower peak intensity of proteomic feature m/z 6626 was significantly associated with higher rate of tumor recurrence **(C)**, and poor overall survival **(D)**.

Table 1
Patient Characteristics

Characteristics	Value
Age, yr	
Median (range)	48 (27–79)
Sex, No. (%)	
Males	47 (73%)
Females	17 (27%)
UICC overall stage, No. (%)	
I	5 (8%)
II	22 (34%)
III	18 (28%)
IV	19 (30%)
Primary tumor (T) stage, No. (%)	
T1	13 (20%)
T2	29 (45%)
T3	12 (19%)
T4	10 (16%)
Regional lymph nodes (N) stage, No. (%)	
N0	22 (34%)
N1	19 (30%)
N2	12 (19%)
N3	11 (17%)
Local Recurrences after RT, No. (%)	14 (22%)
Distant Metastases after RT, No. (%)	13 (20%)

Decision Tree for Predicting Tumor Recurrence

Figure 2 indicates a pruned decision tree resulted from CART modeling for predicting NPC patients with and without tumor recurrences by using the UICC stage and proteomic features m/z 8808 and m/z 6626. The performance of the decision tree model was evaluated by Leave-One-Out cross-validation (Fig. 2). The cross-validation result showed that this decision tree correctly predicted 20 of 24 patients (83%) with local recurrences and/or distant metastases, and 33 of 40 (83%) of patients without recurrence, resulting in an overall accuracy of 83%. This decision tree correctly identified 11 of 13 cases (85%) with distant metastases, and 10 of 14 cases (71%) with local recurrences. Kaplan-Meier analysis showed that the decision tree could predict both recurrence-free survival ($p < 0.00005$; log-rank statistics = 34.1), and overall survival ($p < 0.00005$; log-rank statistics = 27.0) (Fig. 3).

To assess the performance of the decision tree as a prediction tool for important clinical outcomes, we evaluated the positive and negative predictive values for tumor recurrence within 4 yr after achieving complete remission. The positive predictive value was 74%, and the negative predictive value was 89%.

Discussion

About 80% of patients with early-stage (UICC stage I and II) NPC can be cured of their disease with standard radical RT. However in patients with more advanced disease stages (UICC stage III and IV), recurrence is common and, consequently, 5-yr survival rates are only 45–60%. Recent studies have demonstrated an improvement of progression-free survival with concurrent cisplatin and RT compared with RT alone, and the improvement was statistically significant in patients with advanced disease *(12)*. Unfortunately concurrent cisplatin and RT resulted in considerable toxicity. Probably more intensive therapy after concurrent chemoradiation can further improve the treatment results *(14)*, but such strategy will require better selection of patients to confine treatment to those who will benefit most, and to avoid toxic therapy in those who are likely to be already cured.

Previously we reported that level of post-treatment circulating Epstein-Barr virus (EBV) DNA was an independent prognostic indicator for distant metastasis as shown by multivariate Cox-regression analysis *(15)*. However, neither the level of pretreatment EBV DNA nor the level of post-treatment EBV DNA was statistically associated with local recurrence in

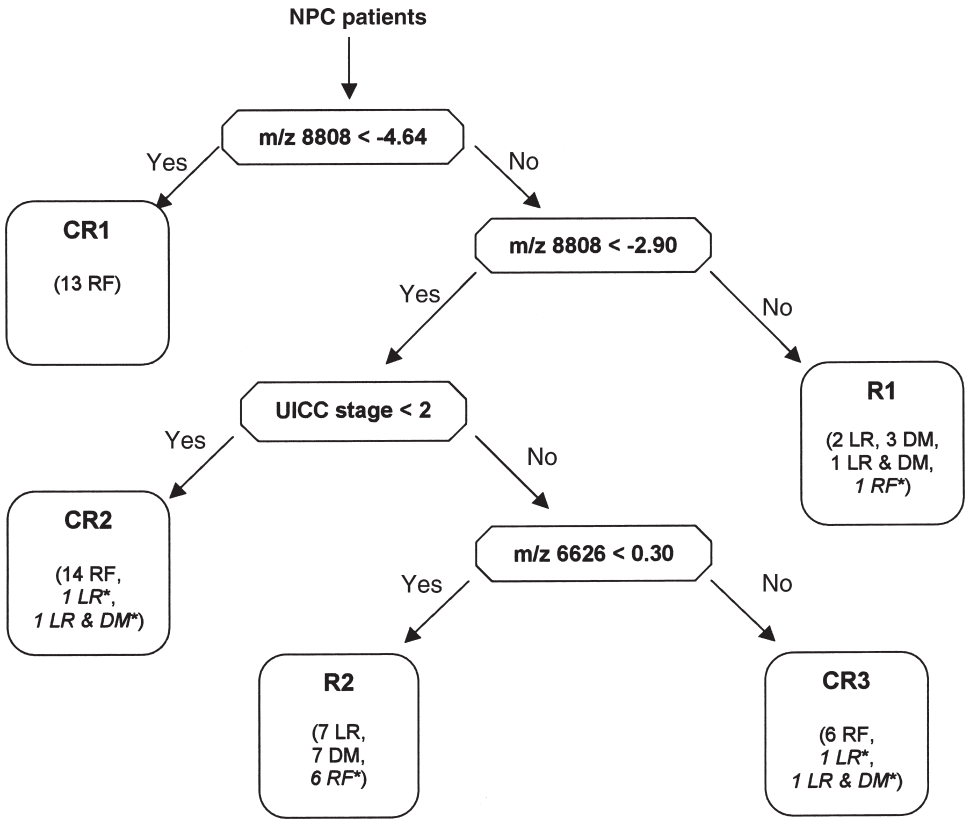

Fig. 2. The pruned decision tree generated from International Union Against Cancer (UICC) overall stage, serum levels (log2 normalized peak intensities) of proteomic feature m/z 8808 and proteomic features m/z 6626 for predicting recurrence after radiotherapy (RT) in nasopharyngeal carcinoma (NPC). The decision tree divides patients into 5 groups. Two groups (R1 and R2) are for the patients with tumor recurrence, whereas three groups (CR1, CR2 and CR3) are for the patients without recurrence in 4 yr after RT. The Leave-One-Out cross-validation result (No.) of the patients without recurrence (recurrence free, RF) and with local recurrence (LR) and/or distant metastasis (DM) is bracketed. The wrongly predicted cases were italic and asterisked (*).

the multivariate analysis. It appears that other serum/plasma markers are needed for predicting local recurrence.

In this study, we demonstrated that a unique proteomic pattern was present in pretreatment sera of the NPC patients who later had local/distant recurrence within 4 yr. Multivariate Cox-regression analysis revealed that two proteomic features with molecular weight of 6626 and 8808 Daltons were statistically significantly associated with the time to first local/distant recurrence. The multivariate analysis result also indicates that these two proteomic features provide prognostic information in additional to the traditional TNM staging system. As inclusion of the proteomic features did not lead to an exclusion of overall UICC stage in the Cox-regression model, this indicates that overall UICC stage provides some prognostic information that the proteomic features do not have. Therefore the two serum proteomic features and overall UICC stage are complementary to each other in predicting tumor recurrence after RT.

Furthermore, the decision tree, based on the peak intensities of these two proteomic features and overall UICC stage, demonstrated high positive and negative predictive values, as

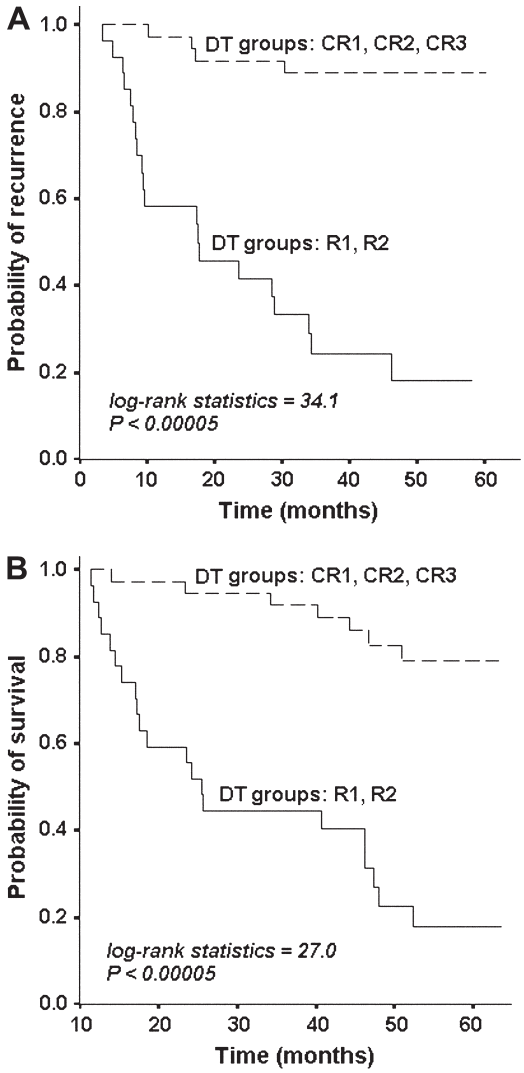

Fig. 3. **(A)** Recurrence-free survival analysis and **(B)** overall survival analysis of the Leave-One-Out cross-validation result of the decision tree (DT) in predicting patients with (DT groups: R1, R2) or without (DT groups: CR1, CR2, CR3) local/distant recurrence after curative radiotherapy.

showed by the Leave-One-Out cross-validation. Although the decision tree could not differentiate local recurrence from distant metastasis, the decision tree correctly predicted patients with local recurrence and patients with distant metastasis at similar sensitivity. Our results suggest that the proteomic pattern comprising the 6626 and 8808 Daltons proteomic features is a potential surrogate marker for predicting not only distant metastasis, but also local recurrence in NPC patients after curative RT. It is important to emphasize that the current study was exploratory in nature. Further studies are needed to validate the use of this predictive model in NPC patients in a prospective manner, or at least to be validated with another retrospective study of larger sample size. It will be also invaluable to investigate whether combined use of circulating EBV DNA and proteomic pattern could further the predictive values for local/distant recurrence, and to investigate the protein identity of these two proteomic features. Nevertheless, because each protein has a unique m/z value, even without knowing their protein identities, they can be unambiguously detected and quantified in patient sera.

Recently, it has been suggested that the findings of studies employing the SELDI ProteinChip technology could be biased by artifacts related to the nature of the clinical samples used, the sample storage conditions, the experimental details, the mass spectrometric instruments, and/or bioinformatic analyses *(16,17)*. In the present study, various preventive measures have been carried out to avoid generation of biased results, as previously described by Poon et al. *(5,6)*. First, all the serum samples were collected and processed within the same clinical and laboratory settings. Second, to ensure the quality of the serum samples, all samples were processed within 1 h after the blood taking, and the serum samples were stored at –70°C before analysis. Third, all samples were analyzed blindly without knowledge of the recurrence status. Fourth, stringent criteria were used to define a proteomic feature as a potential prognostic indicator of tumor recurrence in this study as described in the statistical methods.

In conclusion, the present study demonstrated that a unique serum proteomic pattern is present in the pretreatment sera of NPC patients who subsequently developed local/distant

recurrences following RT. The pattern identified by the SELDI ProteinChip technology together with UICC stage could be used to construct a decision tree in predicting recurrence with high positive and negative predictive values.

Acknowledgments

This work was supported by a Strategic Research Area Grant from the Chinese University of Hong Kong and a Central Allocation Grant from the University Grants Committee of Hong Kong.

References

1. Ho, J.H.C. (1978) An epidemiologic and clinical study of nasopharyngeal carcinoma. Int. J. Radiat. Oncol. Bio. Phys. 4:183–205.
2. Teo, P.M., Yu, P., Lee, W.Y., Leung, S.F., Kwan, W.H., Yu, K.H., et al. (1996) Significant prognosticators after primary radiotherapy in 903 nondisseminated nasopharyngeal carcinoma evaluated by computer tomography. Int. J. Radiat. Oncol. Biol. Phys. 36:291–304.
3. Petricoin III, E.F., Ardekani, A.M., Hitt, A.B., Levine, P.J., Fusaro, V.A., Steinberg, S.M., et al. (2002) Use of proteomic patterns in serum to identify ovarian cancer. Lancet 359:572–577.
4. Poon, T.C.W., Yip, T.T., Chan, A.T.C., Yip, C., Yip, V., Mok, T.S.K., et al. (2003) Comprehensive proteomic profiling identifies serum proteomic signatures for detection of hepatocellular carcinoma and its subtypes. Clin. Chem. 49:752–760.
5. Poon, T.C.W., Chan, K.C.A., Ng, P.C., Chiu, R.W.K., Ang, I.L., Tong, Y.K., et al. (2004) Serial analysis of plasma proteomic signatures in pediatric patients with severe acute respiratory syndrome and correlation with viral load. Clin. Chem. 50:1452–1455.
6. Poon, T.C.W., Hui, A.Y., Chan, H.L.Y., Ang, I.L., Chow, S.M., Wong, N., et al. (2005) Prediction of liver fibrosis and cirrhosis in chronic Hepatitis B Infection by serum proteomic Fingerprinting: a pilot study. Clin Chem. in press.
7. Cho, W.C., Yipm, T.T., Yip, C., Yip, V., Thulasiraman, V., Ngan, R.K.C., et al. (2004) Identification of serum amyloid a protein as a potentially useful biomarker to monitor relapse of nasopharyngeal cancer by serum proteomic profiling. Clin. Cancer Res. 10:43–52.
8. Xiao, Z., Luke, B.T., Izmirlian, G., Umar, A., Lynch, P.M., Phillips, R.K., et al. (2004) Serum proteomic profiles suggest celecoxib-modulated targets and response predictors. Cancer Res. 64:2904–2909.
9. Breiman, L., Friedman, J.H., Olshen, R.A., and Stone, C.J. (1984) Classification and regression trees. In The Wadsworth Statistics/Probability Series. (Bickel, P., Cleveland, W., and Dudley, R., eds.), Belmont: Wadsworth International Group.
10. Poon, T.C.W., Chan, A.T.C., Zee, B., Ho, S.K.W., Mok, T.S.K., Leung, T.W.T., et al. (2001) Application of classification tree and neural network algorithm to the identification of serological liver marker profiles in the diagnosis of hepatocellular carcinoma. Oncology 61:275–283.
11. Markey, M.K., Tourassi, G.D., and Floyd Jr, C.E. (2003) Decision tree classification of proteins identified by mass spectrometry of blood serum samples from people with and without lung cancer. Proteomics 3:1678–1679.
12. Chan, A.T.C., Teo, P.M.L., Nagan, R.K., Leung, T.W., Lau, W.H., Zee, B., et al. (2002) Concurrent chemotherapy-radiotherapy compared with radiotherapy alone in locoregionally advanced nasopharyngeal carcinoma: progression free survival analysis of a phase III randomized trial. J. Clin. Oncol. 20:2038–2044.
13. Radmacher, M.D., McShane, L.M., and Simon, R. (2002) A paradigm for class prediction using gene expression profiles. J. Comput. Biol. 9:505–511.
14. Al-Sarraf, M., LeBlanc, M., Giri, P.G., Fu, K.K., Cooper, J., Vuong, T., et al. (1998) Chemoradiotherapy versus radiotherapy in patients with advanced nasopharyngeal cancer: phase III randomized intergroup study 0099. J. Clin. Oncol. 16:1310–1317.
15. Chan, A.T.C., Lo, Y.M.D., Zee, B., Chan, L.Y.S., Ma, B.B.Y., Leung, S.F., et al. (2002) Plasma Epstein-Barr virus DNA and residual disease after radiotherapy after radiotherapy for undifferentiated carcinoma. J. Natl. Cancer. Inst. 94:1614–1619.
16. Diamandis, E.P. (2003) Proteomic patterns in biological fluids: do they represent the future of diagnostics? Clin. Chem. 49:1272–1275.
17. Diamandis, E.P. (2004) Analysis of serum proteomic patterns for early cancer diagnosis: drawing attention to potential problems. J. Natl. Cancer Inst. 96:353–256.