Nucl Med Mol Imaging
Vol. 41, No. 4, Aug 2007

Original Article

의사결정트리 프로그램 개발 및 갑상선유두암에서 질량분석법을 이용한 단백질 패턴 분석

아주대학교 의과대학 핵의학과학교실, 건국대학교 컴퓨터공학과[1]
윤준기 · 이 준[1] · 안영실 · 박복남 · 윤석남

Development of Decision Tree Software and Protein Profiling using Surface Enhanced Laser Desorption/Ionization - Time of Flight - Mass Spectrometry (SELDI-TOF-MS) in Papillary Thyroid Cancer

Joon-Kee Yoon, M.D., Jun Lee, B.S.[1], Young-Sil An, M.D.,
Bok-Nam Park, M.S., and Seok Nam Yoon, M.D.

Department of Nuclear Medicine & Molecular Imaging, Ajou University School of Medicine, Suwon and [1]Department of Computer Engineering, Konkuk University, Seoul, Korea.

Purpose: The aim of this study was to develop a bioinformatics software and to test it in serum samples of papillary thyroid cancer using mass spectrometry (SELDI-TOF-MS). **Materials and Methods:** Development of 'Protein analysis' software performing decision tree analysis was done by customizing C4.5. Sixty-one serum samples from 27 papillary thyroid cancer, 17 autoimmune thyroiditis, 17 controls were applied to 2 types of protein chips, CM10 (weak cation exchange) and IMAC3 (metal binding - Cu). Mass spectrometry was performed to reveal the protein expression profiles. Decision trees were generated using 'Protein analysis' software, and automatically detected biomarker candidates. Validation analysis was performed for CM10 chip by random sampling. **Results:** Decision tree software, which can perform training and validation from profiling data, was developed. For CM10 and IMAC3 chips, 23 of 113 and 8 of 41 protein peaks were significantly different among 3 groups ($p<0.05$), respectively. Decision tree correctly classified 3 groups with an error rate of 3.3% for CM10 and 2.0% for IMAC3, and 4 and 7 biomarker candidates were detected respectively. In 2 group comparisons, all cancer samples were correctly discriminated from non-cancer samples (error rate = 0%) for CM10 by single node and for IMAC3 by multiple nodes. Validation results from 5 test sets revealed SELDI-TOF-MS and decision tree correctly differentiated cancers from non-cancers (54/55, 98%), while predictability was moderate in 3 group classification (36/55, 65%). **Conclusion:** Our in-house software was able to successfully build decision trees and detect biomarker candidates, therefore it could be useful for biomarker discovery and clinical follow up of papillary thyroid cancer. (Nucl Med Mol Imaging 2007;41(4):299-308)

Key Words: biomarker discovery, SELDI-TOF-MS, decision tree, papillary thyroid cancer

Introduction

Postoperative high dose radioiodine therapy with ^{131}I is strongly recommended in the treatment of differentiated thyroid cancer in that it provides better visualization of recurrent disease by ablating thyroid remnant. After ablation of thyroid bed, ^{131}I whole body scintigraphy and serum thyroglobulin are generally used as follow-up methods. However, whole body scan in this clinical situation has very low diagnostic accuracy and a disadvantage of possible stunning effect decreasing the

• **Received**: 2007. 3. 30. • **Accepted**: 2007. 7. 8.
• Address for reprints: Seok Nam Yoon, M.D., Department of Nuclear Medicine & Molecular Imaging, Ajou University, School of Medicine, San 5, Wonchun-dong, Youngtong-gu, Suwon, Kyunggi-do, Korea, 721-443
Tel: 82-31-219-5948, Fax: 82-31-219-5950
E-mail: snyoon@ajou.ac.kr
※This study was supported by a grant from Korean society of nuclear medicine (2004).
※We appreciate professor Won Yonggwan (Chonnam University)'s comments for developing this decision tree software and Park Seungpil (BMS Korea)'s technical assistance for SELDI-TOF-MS.

efficacy of subsequent therapeutic dose.[1,2] For this reason, alternative methods such as serum thyroglobulin with non-radioisotope scanning or ^{123}I scanning have been proposed for the follow-up methods of thyroid cancer. Serum thyroglobulin monitoring has also a limitation of insensitivity in the presence of circulating anti-thyroglobulin autoantibody, which has not been overcome yet. In addition, both tests require a long-term withdrawal of thyroxine medication inducing intolerable hypothyroid symptoms, which sometimes are the reason for refusing those tests. Although head and neck ultrasonography, chest X-ray, computed tomography and positron emission tomography may play ancillary roles, a more comfortable and reliable test is needed for differentiated thyroid cancers.

Surface Enhanced Laser Desorption/Ionization - Time of flight - Mass spectrometry (SELDI-TOF-MS) is a recently developed, high throughput proteomic technique which identifies disease-related proteins (biomarkers) using various protein chip arrays, a protein chip reader and bioinformatics softwares. Like other MS techniques, SELDI-TOF-MS is more sensitive in low molecular weight proteins (< 20 kDa) than two dimensional polyacrylamide gel electrophoresis, thus can play a complementary role in proteomic research. Besides, it can extract protein profiles from as few as 25-50 cells, and various types of biological fluids (serum, urine, blood, plasma, cerebrospinal fluid, cell lysates, etc.) can be applied.[3,4] After protein chips are loaded on its reader, they are irradiated and give rise to gaseous ion by desorption and ionization processes. Liberate gaseous ions fly through the ion chamber, and the detector measures their *m/z* ratios to generate protein profiles through baseline subtraction, peak detection and peak alignment.[4,5] For further clinical application or biomarker discovery, a decision tree software, which can elucidate differences of protein expression pattern between diseased samples and controls, is needed.

Recently, SELDI-TOF-MS is increasingly used for discovering biomarkers of various cancers and neurologic diseases such as ovarian cancer, breast cancer, prostate cancer and Alzheimer's disease. However, there has been no effort to develop a new diagnostic test using SELDI-TOF-MS in the field of nuclear medicine as well as differentiated thyroid cancer.[6] The aim of this study was to develop a decision tree generating software, named 'Protein-analysis', by customizing C4.5 (by J. R. Quinlan, 1993) and to test this software with serum samples from patients with papillary thyroid cancer.[7-9]

Materials and Methods

1. Development of Decision Tree Software

By customizing C4.5, we developed a 'Protein-analysis' software performing a binary search to generate a decision tree.[7-9] Like other decision trees, it had a flow-chart like tree structure and consisted of sequential questions, 'node'. Each parent node, by which samples were divided into 2 subgroups (ex. cancer or control) based on *m/z* ratio and its intensity (threshold), was connected to 2 children nodes. In a decision tree, the top node was named as 'root node' followed by sequential children nodes named 'non-terminal nodes' and 'terminal nodes' thereafter. When nodes satisfied stopping criterion: subgroup with single class or subgroup with minimal error, those nodes were determined as 'terminal nodes'. Summed error rate of all terminal nodes was used to evaluate the accuracy of SEDLI-TOF-MS analysis and decision trees.

2. Collection and storage of samples

A total of 61 serum samples of 27 patients with papillary thyroid cancer, 17 with autoimmune thyroiditis, 17 controls were consecutively collected for 5 months and stored at -70℃. All samples from cancer patients were acquired 1 day before thyroidectomy. The presence of autoantibody in patients with clinically suspected autoimmune thyroiditis was confirmed by RIA. All samples from healthy controls were tested for blood count, serum glucose concentration, liver function test, electrolytes, thyroid function test, urine analysis and hepatitis antigen/antibody test, which revealed no evidence of cancer, inflammatory disease or other severe medical problems. Any hemolytic sample was excluded from the study to avoid artifactual formation of hemoglobin peak.

Two types of protein chips, CM10 (weak cation exchange) and IMAC3 (immobilized metal affinity capture - Cu), were used, because those chips were proved to

detect more protein peaks in serum samples.[4,5] For CM10 chip, 60 samples were applied (all samples except 1 cancer sample), while for IMAC3 chip, 50 samples were used (all 27 cancer, 13 autoimmune thyroiditis and 10 control samples).

3. Preparation of samples for protein profiling

All samples were thawed simultaneously at room temperature, and 100 μL of each sample was centrifuged at 4℃, 14,000 rpm for 30 min for delipidation. Urea lysis buffer (30 μL, 8M urea with 1% CHAPS) and 150 μL binding buffer (50 mM sodium acetate with 0.1% trixon X-100 for CM10, pH = 5.2; 150 mM NaCl/PBS buffer for IMAC3, pH = 7.0) were mixed with 20 μL delipidated serum for 10 min.

4. Loading of samples

CM10

CM10 chips were pretreated with 10 μL, 10 mM HCl for 10 min and rinsed briefly with 10 mL deionized water for 3 times. And after assembling a bioprocessor (Ciphergen Biosystems, Fremont, CA, USA), 350 μL binding buffer was added to each well and shaken vigorously for 5 min, 250 rpm, at room temperature. After removing binding buffer from each well, 100 μL diluted serum sample was added to each well and incubated with vigorous shaking for 30 min. And samples were removed and each well was washed twice with 350 μL binding buffer for 5 min. Chips were removed from the bioprocessor and rinsed briefly with 8 mL deionized water in a 15 mL conical tube, then 1 μL EAM (energy absorbing material) solution was applied twice on each spot of air-dried chips.

IMAC3

After assembling a bioprocessor, IMAC3 chips were pretreated twice with 10 μL, 100 mM copper sulfate for 15 min with shaking, and rinsed briefly with 50 μL deionized water twice. And each spot was rinsed briefly with 50 μL, 50 mM sodium acetate buffer (pH=4) and 50 μL deionized water (twice). Then, 300 μL, 150 mM NaCl buffer was added to each well and shaken vigorously for 5 min, 250 rpm, at room temperature. After removing binding buffer from each well, 100 μL diluted serum sample was added to each well and incubated with vigorous shaking for 30 min. Remaining procedures were the same with those for CM10.

Protein profiling with SELDI-TOF-MS

Mass spectrometry analysis of serum samples was performed using Proteinchip Reader (Ciphergen Biosystems, Fremont, CA, USA). Mass spectra of all samples were normalized by baseline subtraction. Multiple protein peaks were detected automatically, then adjacent peaks (mass error score $\leq$ 0.18%) were recognized as the same peak by peak alignment process. Protein profiling data including mass-to-charge (m/z) ratio and peak intensity were exported to excel files. Among all spectra, protein peaks with m/z ratio from 3,000 to 30,000 Da were selected, because the accuracy of SELDI TOF MS is relatively low outside this range. Exported excel files were applied to the 'Protein analysis' software. Decision tree analyses were performed among all 3 groups (control, thyroiditis and cancer) as well as between 2 groups (control vs. cancer and thyroiditis vs. cancer).

Validation of SELDI assisted disease classification using Decision Tree software

To validate the SELDI assisted classification in papillary thyroid cancer, we randomly divided the samples, analyzed by CM10 chip, into 2 sets; training and test set. With training set, the classifiers (nodes) were determined and its informations were saved as files, then it was validated in test set. Among 60 samples, 3 controls, 3 autoimmune thyroiditis, and 5 papillary thyroid cancers were randomly selected as test set, and the remainders (14 controls, 14 autoimmune thyroiditis, and 21 papillary thyroid cancers) were classified as training set. Five tests were performed and summed predictability was calculated. With IMAC3 chip, validation was not performed because of its small sample number.

Results

1. Decision Tree software

'Protein analysis' software had 2 major functions, 'training' and 'validation', and 3 display windows. 'Training' function included 'cumulative training' to accumulate

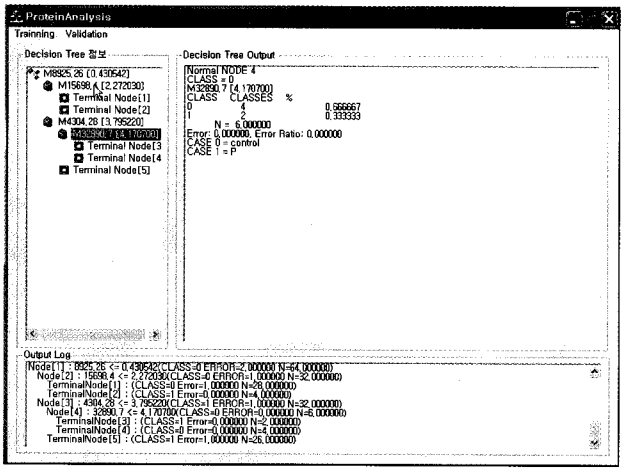

Fig. 1. 'Protein analysis' software consists of 3 windows displaying tree structure (left), outputs of nodes (right), and tree building processes and validation results (bottom).

proteomic data analyzed at different time points and build a large database. Left window displayed tree structure and right window showed outputs of each node. Tree building processes and results of validation were displayed in the bottom window (Fig. 1).

2. Development and validation of decision tree

For CM10 chip, 113 protein peaks were detected by SELDI-TOF-MS analysis, and among those, intensities of 23 peaks were significantly different between 3 groups ($p < 0.05$). In 2 group comparisons, 27 peaks of controls and 6 of thyroiditis samples were significantly different from those of cancer. For IMAC3 chip, a total of 41 peaks were detected. The intensities of 8 peaks were significantly different among 3 groups, and 7 peaks of controls and 6 of thyroiditis samples were significantly different from those of cancer samples (Fig. 2). Decision tree correctly classified 3 groups with an error rate of 3.3% for CM10 and 2.0% for IMAC3, and 4 and 7 biomarker candidates were detected respectively (Fig. 3).

In 2 group comparisons for CM10 chips, single node correctly differentiated all cancer samples from controls or autoimmune thyroiditis samples (Figure 4, error rate = 0).

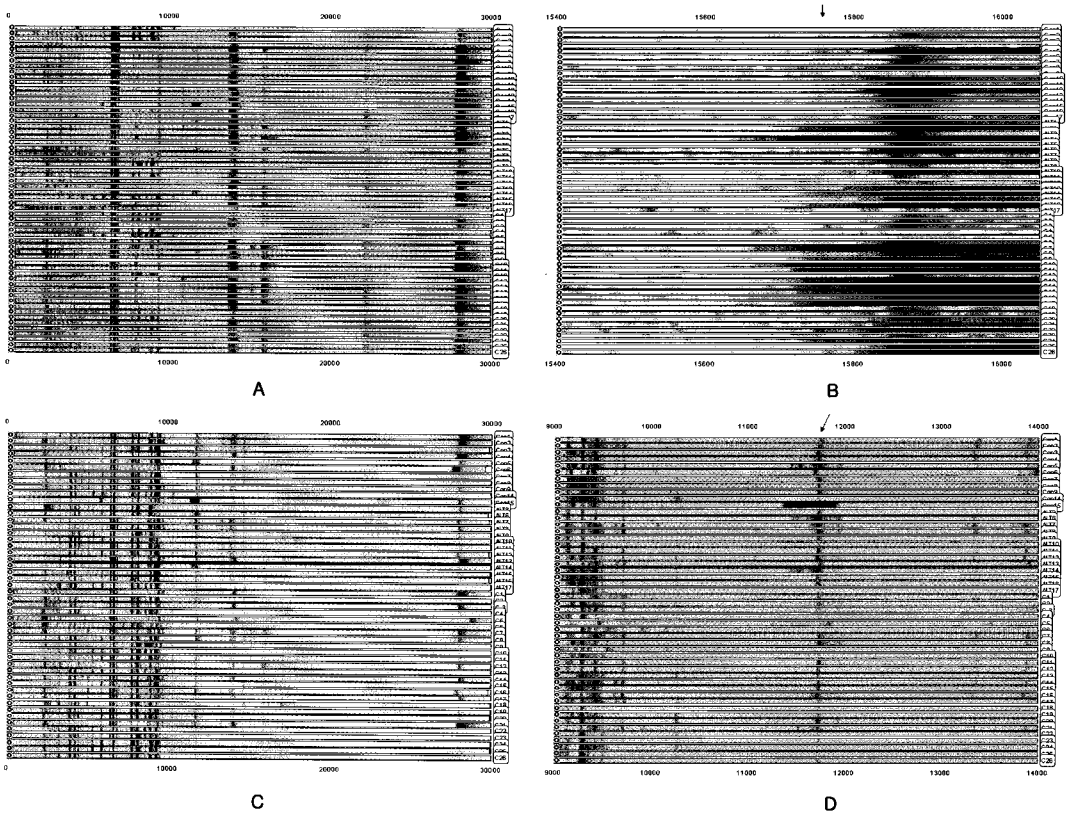

Fig. 2. Full spectrum of mass analysis with CM10 chips (A) and IMAC3 chips (C) are displayed as gel view. B and D are showing representative nodes for CM10 (15.7 kDa, B) and IMAC3 (11.7 kDa, D) (arrow).

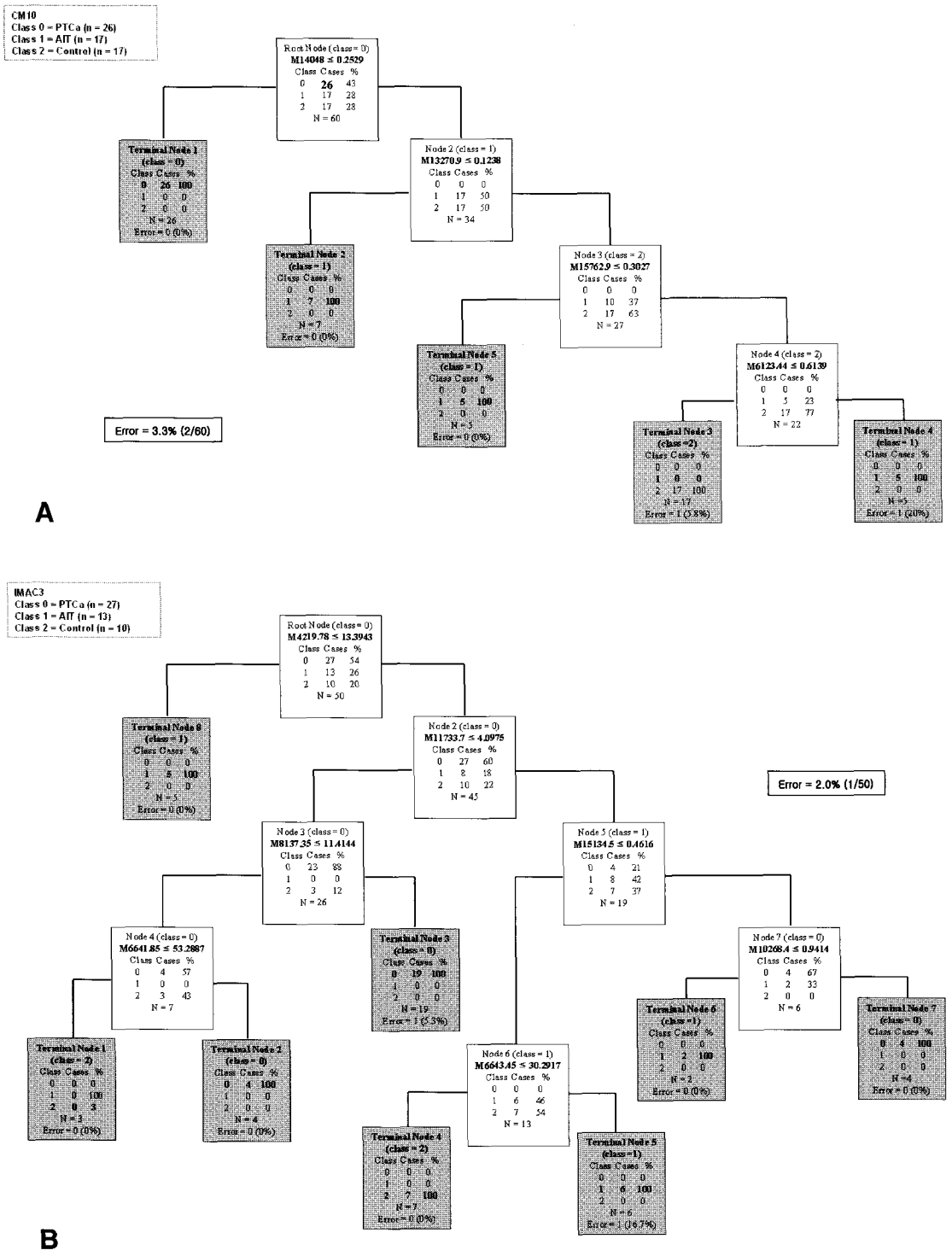

Fig. 3. Decision tree analysis and CM10 protein chips correctly classified 60 samples into 3 groups using 4 nodes (14.0, 13.2, 15.8 and 6.1 kDa, error rate = 3.3%) (A). Decision tree analysis and IMC3 chips also correctly classified 50 samples into 3 groups using 7 nodes (4.2, 11.7, 8.1, 6.6, 15.1, 6.6 and 10.3 kDa, error rate = 2.0%) (B); PTCa = papillary thyroid cancer, AIT = autoimmune thyroiditis.

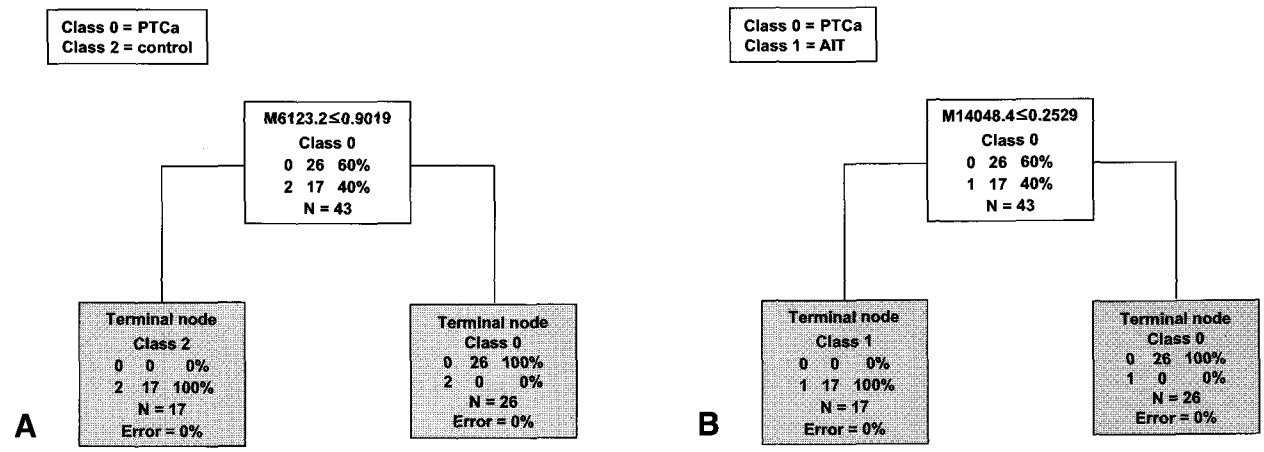

Fig. 4. Decision tree completely differentiated cancer samples from controls (A) and autoimmune thyroiditis (B) samples for CM10 chip. Single node (at 6.1 kDa and 14.0 kDa) was used for each classification; PTCa = papillary thyroid cancer, AIT = autoimmune thyroiditis.

Table 1. Validation Test for Mass Analysis Using CM10 Chips

*Samples	Validation 1	Validation 2	Validation 3	Validation 4	Validation 5
c	a	c	a	c	a
c	c	a	a	a	a
c	a	c	c	c	c
a	c	a	c	c	a
a	c	a	a	c	c
a	c	a	c	c	c
p	c	p	p	p	p
p	p	p	p	p	p
p	p	p	p	p	p
p	p	p	p	p	p
p	p	p	p	p	p
†3G	6/11	1/11	4/11	4/11	4/11
‡2G	1/11	0/11	0/11	0/11	0/11

*c; control, a: autoimmune thyroiditis, p; papillary thyroid cancer
†3G: error rate in differentiating 3 groups
‡2G: error rate in differentiating cancers from others (control and thyroiditis)

As for IMAC3, cancer samples were correctly discriminated from controls and autoimmune thyroiditis samples by multiple nodes (Fig. 5, error rate = 0). For CM10 chip, all nodes used for 2-group comparisons were the classifiers of 3-group comparison, whereas, for IMC3 chip, some nodes of 2-group comparison (at 14 kDa and 6.4 kDa) were not the classifiers of 3-groups comparison.

Validation results from 5 different test sets revealed SELDI-TOF-MS and decision tree analysis correctly predicted (54/55, 98%) unknown samples in differentiating cancers from non-cancer samples, while predictability was moderate (36/55, 65%) in 3 group classification (Table 1).

Discussions

In this study, we developed a decision tree software for biomarker discovery, and tested in serum samples of papillary thyroid cancer with 2 types of protein chips. And as a result, our decision tree software successfully differentiated cancer samples from controls and autoimmune thyroiditis samples and detected a few biomarker candidates.

To develop a decision tree software, we used C4.5 designed by J. R. Quinlan.[7,9,10] Like CART (classification and regression tree), C4.5 is a machine learning algorithm which can build a small and accurate decision tree. Although a few decision tree softwares derived from C4.5

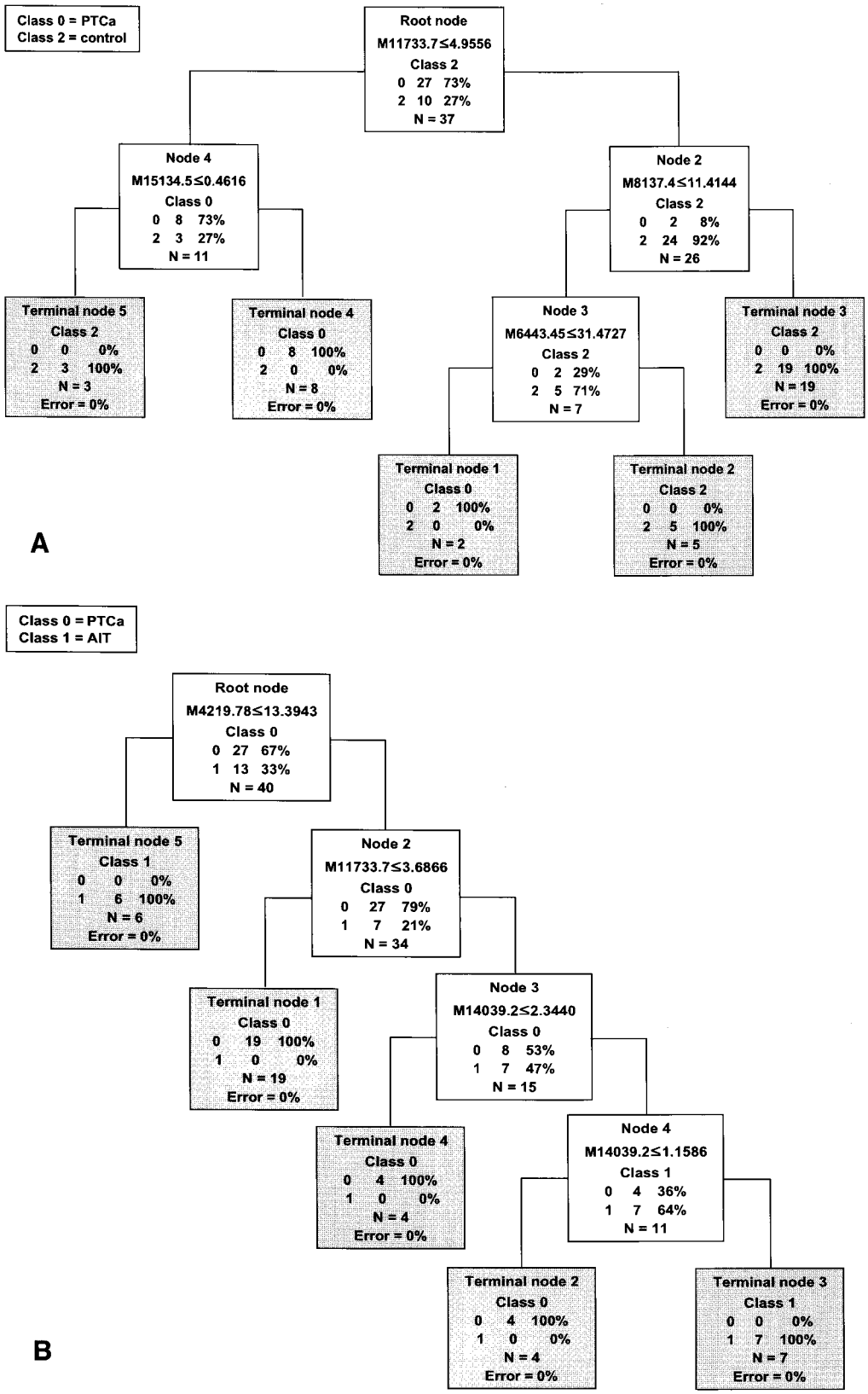

Fig. 5. Decision tree completely differentiated cancer samples from controls (A) and autoimmune thyroiditis (B) samples for IMAC3 chip. Some nodes were not the classifiers selected for 3 groups comparison; PTCa = papillary thyroid cancer, AIT = autoimmune thyroiditis.

or CART are commercially available (as an example, CART® by Salford system or Biomarker pattern software® by Ciphergen Biosystems), those are very expensive and those functions are discordant with our purpose. We intended to apply our decision trees not only to biomarker discovery but to clinical diagnosis. For the latter purpose, a database of protein profiles should be large enough to accurately classify a single sample into either 'disease' or 'healthy'. Meanwhile, for the former purpose, a threshold for peak alignment (0.3% in this study) should be changeable. Therefore, we developed our own software for this study.

Other than protein biomarker discovery, C4.5 and CART have been used in clinical decision making and prediction in the fields of clinical oncology, pathology, cardiology, traumatic injury, etc.[11-24] It is probably due to its wide acceptability for various clinical parameters such as clinical symptoms, blood tests, EKG, histologic results etc. For this purpose, CART seems more popular than C4.5.

A decision tree consists of numerous nodes, which classify samples into 2 or more subgroups according to a criterion based on mass and peak intensity of proteins. Decision tree algorithm selects mass and intensity threshold using cost function to make subgroups become as pure as possible.[8] When subgroups are indivisible, they are called 'terminal nodes'. The sum of heterogeneity of terminal nodes is a error rate of decision tree analysis, and some of non-terminal nodes (including root node) can be biomarker candidates. As an example (Fig. 3), in 3 group-decision tree analysis with CM10 chips, 1 root node (14.0 kDa $\leq$ 0.2529) and 3 non-terminal nodes (13.3 kDa $\leq$ 0.1238, 15.8 kDa $\leq$ 0.3027 and 6.1 kDa $\leq$ 0.6139) can be biomarker candidates. Root node means that any sample with 14.0 kDa protein of its intensity less than 0.2529 can be classified as cancer. Considering error rate (0%) of terminal node 1, root node is a good biomarker candidate for papillary thyroid cancer, and worthy of identification. The other non-terminal nodes also can be biomarker candidates for discriminating autoimmune thyroiditis samples from controls. For biomarker discovery, a decision tree with smaller number of non-terminal nodes and low error rate is more desirable. In this study, decision tree analysis with CM10 chip used 4 non-terminal nodes for 3-group comparison and single node for 2 group comparisons, whereas, those with IMAC3 employed 7, 4 and 4 non-terminal nodes for the same comparisons (Fig. 3-5). If the samples for both tests were identical, CM10 chip would be superior to IMAC3 for biomarker discovery of papillary thyroid cancer.

In this work, because of a small sample size, validation test was performed by random allocation method within the same group. Validation result with CM10 chip was excellent for differentiation of cancer from non-cancer samples, whereas it was moderate for 3 group discrimination. Therefore, in clinical practice, a decision tree analysis could be used as a follow up method of papillary thyroid cancer after radioiodine ablation, but not for a screening method.

Protein chips have specific chromatographic surfaces with broad binding properties (hydrophobic, hydrophilic, anionic, cationic or immobilized metal affinity capture) or with covalently-bound, specific molecules (antibodies, receptors, enzymes, DNA, small molecules, ligands and lectins).[4,5,25] In this study, we used weak cation exchange protein chips, CM10, and immobilized metal affinity capture protein chips, IMAC3, because these chips are generally used in differentiating diseased samples from controls due to its broad binding properties. Pilot profiling is occasionally recommended for determinating the most suitable chip.[4,8] It is usually performed with 4 types of broad binding chips (hydrophobic, anion exchange, cation exchange and metal chips). Because one chip has 8 spots, maximum 4 controls and 4 disease samples can be applied to each chip. The chip showing more disease-specific protein peaks than others used to be selected.

Biomarker discovery using SELDI-TOF-MS has been used in the field of cancer, infection, neurologic disease and cardiovascular disease.[5] Early attempts for tumor marker detection were focused on prostate, breast and ovarian cancers.[26-35] Although several biomarkers were found and validated for those cancers, cross validations between laboratories or between protocols are needed before identification of biomarkers.[5,36,37] Baggerly[36] demonstrated that an unrecognized change between protocols could make a shift of protein peaks obtained from serum samples. Bons[37] also found that the repeated thaw-freeze cycle

could make artifactual peaks which were significantly different between controls and cancer patients. In addition, sampling time, storage method, fractionation method and patients' status (fasting, age, etc) also should be controlled to minimize the error in preparing samples. Another problem with SELDI-assisted biomarker discovery is whether identified proteins are cancer-specific. While non-cancer proteins or non-specific cancer-associated proteins also could be used for discriminating cancer samples from controls, those proteins can not be used as biomarkers of specific cancers. Therefore, reproducibility and cancer-specificity are questions to be solved before development of a new clinical test. In this study, we performed a validation of biomarker candidates within the same samples using random allocation. For further development of a new diagnostic test for thyroid cancer, intra- and inter-laboratory cross-validations for reproducibility and tests for cancer-specificity should be performed.

As a conclusion, our in-house software 'Protein analysis' was successfully developed and built decision trees from protein profiling data produced by SELDI-TOF-MS. It also detected a few biomarker candidates and validated them. Therefore, our decision tree software and accumulated serum protein profiles could be useful for biomarker discovery and clinical follow up of papillary thyroid cancer.

요 약

목적: 본 연구의 목적은 의사결정트리를 생성하는 생물정보학 프로그램을 개발하고, 이를 갑상선유두암 혈청의 질량분석자료로 시험해 보는 것이다. **대상 및 방법:** C4.5를 커스터마이징하여 의사결정트리 분석을 수행할 수 있는 'Protein analysis'라는 프로그램을 개발하였다. 61개의 혈청 시료(갑상선유두암 27, 자가면역성 갑상선염 17, 대조군 17)를 일정 기간 동안 순차적으로 냉동한 후 실온에서 일시에 해동하여 분석에 사용하였다. 모든 시료는 탈지질화 과정을 거쳐 준비한 후, 2종류의 단백질칩(CM10, IMAC3)에 각각 60개, 50개 시료를 적용하였다. 갑상선유두암의 특징적인 단백질 패턴을 찾기 위해 질량분석기를 이용하여 단백질칩을 분석했다. 'Protein analysis' 프로그램을 이용하여 단백질분포 자료로부터 의사결정트리를 작성하고, 생체표지자 후보물질을 검출하였다. CM10칩에서 발견된 생체표지자 후보물질을 무작위 표본추출 방법을 이용하여 검증하였다. **결과:** 단백질분포 자료의 훈련과 검증이 가능한 의사결정트리 프로그램이 개발되었으며, 이 프로그램은 트리 구조와 노드 정보, 트리 구성 과정을 표시하는 3개의 창으로 구성되었다. CM10칩을 이용한 분석에서 총 113개의 단백질 피크 중 23개가 3그룹 간에 유의한 차이가 있었으며, IMAC3는 41개의 단백질 피크 중 8개가 3그룹 간에 유의한 차이가 있었다. 3그룹 분석에서 의사결정트리는 CM10칩과 IMAC3의 단백질 분포 자료로부터 각각 60개와 50개의 시료를 높은 정확도로 분류하였으며(오차율 = 각각 3.3%, 2.0%), 각각 4개와 7개의 생체표지자 후보물질을 검출하였다. 암시료와 비암시료를 구분하는 2그룹 분석에서, 의사결정트리는 모든 암시료를 정확히 구분하였으며(모두 오차율 = 0%), CM10칩을 이용한 분석에서는 단일 노드를 사용하고, IMAC3칩을 이용한 분석에서는 여러 개의 노드를 사용하였다. CM10칩의 단백질 분포자료를 5번의 무작위 추출에 의해 시행한 검증에서, 암시료와 비암시료를 구분하는데 높은 정확도를 보였으나(정확도 = 98%, 54/55), 3그룹을 구분할 때는 중등도의 정확도를 보였다(정확도 = 65%, 36/55). **결론:** 우리가 개발한 프로그램은 질량분석 자료로부터 성공적으로 의사결정트리를 생성하고, 생체표지자 후보물질을 검출할 수 있었다. 따라서 이 프로그램은 혈청 시료를 이용한 생체표지자 발굴 및 갑상선유두암의 추적관찰에 유용하게 사용될 수 있을 것이다.

References

1. Mansi L, Moncayo R, Cuccurullo V, Dottorini ME, Rambaldi PF. Nuclear medicine in diagnosis, staging and follow-up of thyroid cancer. ***Q J Nucl Med Mol Imaging*** 2004;48:82-95.
2. Ringel MD, Ladenson PW. Controversies in the follow-up and management of well-differentiated thyroid cancer. ***Endocr Relat Cancer*** 2004;11:97-116.
3. Xiao Z, Prieto D, Conrads TP, Veenstra TD, Issaq HJ. Proteomic patterns: their potential for disease diagnosis. ***Mol Cell Endocrinol*** 2005;230:95-106.
4. Issaq HJ, Veenstra TD, Conrads TP, Felschow D. The SELDI-TOF MS approach to proteomics: protein profiling and biomarker identification. ***Biochem Biophys Res Commun*** 2002;292:587-92.
5. Engwegen JY, Gast MC, Schellens JH, Beijnen JH. Clinical proteomics: searching for better tumour markers with SELDI-TOF mass spectrometry. ***Trends Pharmacol Sci*** 2006;27:251-9.
6. Haberkorn U, Altmann A, Eisenhut M. Functional genomics and proteomics-the role of nuclear medicine. ***Eur J Nucl Med Mol Imaging*** 2002;29:115-32.
7. Quinlan JR. Learning logical definitions from relations. ***Machine Learning*** 1990;5:239-66.
8. Won Y, Song HJ, Kang TW, Kim JJ, Han BD, Lee SW. Pattern analysis of serum proteome distinguishes renal cell carcinoma from

other urologic diseases and healthy persons. ***Proteomics*** 2003;3:2310-6.
9. Salzberg SL. Book Review: C4.5: Programs for Machine Learning by J. Ross Quinlan. Morgan Kaufmann Publishers, Inc.,1993. ***Machine Learning*** 1994;16:235-40.
10. Quinlan JR. Improved use of continuous attributes in C4.5. ***J Artif Intell Res*** 1996;5:139-61.
11. Masic N, Gagro A, Rabatic S, Sabioncello A, Dasic G, Jaksic B, et al. Decision-tree approach to the immunophenotype-based prognosis of the B-cell chronic lymphocytic leukemia. ***Am J Hematol*** 1998;59:143-8.
12. Twa MD, Parthasarathy S, Roberts C, Mahmoud AM, Raasch TW, Bullimore MA. Automated decision tree classification of corneal shape. ***Optom Vis Sci*** 2005;82:1038-46.
13. Chan AL, Chen JX, Wang HY. Application of data mining to predict the dosage of vancomycin as an outcome variable in a teaching hospital population. ***Int J Clin Pharmacol Ther*** 2006; 44:533-8.
14. van Gerven MA, Jurgelenaite R, Taal BG, Heskes T, Lucas PJ. Predicting carcinoid heart disease with the noisy-threshold classifier. ***Artif Intell Med*** 2006 (epub).
15. Esquerda A, Trujillano J, Lopez UI, Bielsa S, Madronero AB, Porcel JM. Classification tree analysis for the discrimination of pleural exudates and transudates. ***Clin Chem Lab Med*** 2007;45: 82-7.
16. Takahashi O, Cook EF, Nakamura T, Saito J, Ikawa F, Fukui T. Risk stratification for in-hospital mortality in spontaneous intracerebral haemorrhage: a Classification and Regression Tree analysis. ***Qjm*** 2006;99:743-50.
17. Suzuyama K, Shiraishi T, Oishi T, Ueda S, Okamoto H, Furuta M, et al. Combined proteomic approach with SELDI-TOF-MS and peptide mass fingerprinting identified the rapid increase of monomeric transthyretin in rat cerebrospinal fluid after transient focal cerebral ischemia. ***Brain Res Mol Brain Res*** 2004;129:44-53.
18. Chang LY, Wang HW. Analysis of traffic injury severity: an application of non-parametric classification tree techniques. ***Accid Anal Prev*** 2006;38:1019-27.
19. Webster AP, Goodacre S, Walker D, Burke D. How do clinical features help identify paediatric patients with fractures following blunt wrist trauma? ***Emerg Med J*** 2006;23:354-7.
20. Spurgeon SE, Hsieh YC, Rivadinera A, Beer TM, Mori M, Garzotto M. Classification and regression tree analysis for the prediction of aggressive prostate cancer on biopsy. ***J Urol*** 2006; 175:918-22.
21. Isbister GK, Sibbritt D. Developing a decision tree algorithm for the diagnosis of suspected spider bites. ***Emerg Med Australas*** 2004;16:161-6.
22. Lehmann G, Schmitt C, Kehl V, Schmieder S, Schomig A. Electrocardiographic algorithm for assignment of occluded vessel in acute myocardial infarction. ***Int J Cardiol*** 2003;89:79-85.
23. Langlois RG, Trebes JE, Dalmasso EA, Ying Y, Davies RW, Curzi MP, et al. Serum protein profile alterations in hemodialysis patients. ***Am J Nephrol*** 2004;24:268-74.
24. Smolle J, Kahofer P. Automated detection of connective tissue by tissue counter analysis and classification and regression trees. ***Anal Cell Pathol*** 2001;23:153-8.
25. Merchant M, Weinberger SR. Recent advancements in surface-enhanced laser desorption/ionization-time of flight-mass spectrometry. ***Electrophoresis*** 2000;21:1164-77.
26. Jr GW, Cazares LH, Leung SM, Nasim S, Adam BL, Yip TT, et al. Proteinchip(R) surface enhanced laser desorption/ionization (SELDI) mass spectrometry: a novel protein biochip technology for detection of prostate cancer biomarkers in complex protein mixtures. ***Prostate Cancer Prostatic Dis*** 1999;2:264-76.
27. Paweletz CP, Trock B, Pennanen M, Tsangaris T, Magnant C, Liotta LA, et al. Proteomic patterns of nipple aspirate fluids obtained by SELDI-TOF: potential for new biomarkers to aid in the diagnosis of breast cancer. ***Dis Markers*** 2001;17:301-7.
28. Li J, Zhang Z, Rosenzweig J, Wang YY, Chan DW. Proteomics and bioinformatics approaches for identification of serum biomarkers to detect breast cancer. ***Clin Chem*** 2002;48:1296-304.
29. Petricoin EF, Ardekani AM, Hitt BA, Levine PJ, Fusaro VA, Steinberg SM et al. Use of proteomic patterns in serum to identify ovarian cancer. ***Lancet*** 2002;359:572-7.
30. Streckfus CF, Bigler LR, Zwick M. The use of surface-enhanced laser desorption/ionization time-of-flight mass spectrometry to detect putative breast cancer markers in saliva: a feasibility study. ***J Oral Pathol Med*** 2006;35:292-300.
31. Qu Y, Adam BL, Yasui Y, Ward MD, Cazares LH, Schellhammer PF, et al. Boosted decision tree analysis of surface-enhanced laser desorption/ionization mass spectral serum profiles discriminates prostate cancer from noncancer patients. ***Clin Chem*** 2002;48: 1835-43.
32. Grizzle WE, Adam BL, Bigbee WL, Conrads TP, Carroll C, Feng Z, et al. Serum protein expression profiling for cancer detection: validation of a SELDI-based approach for prostate cancer. ***Dis Markers*** 2003;19:185-95.
33. Vlahou A, Laronga C, Wilson L, Gregory B, Fournier K, McGaughey D, et al. A novel approach toward development of a rapid blood test for breast cancer. ***Clin Breast Cancer*** 2003;4: 203-9.
34. Zhang Z, Bast RC, Jr., Yu Y, Li J, Sokoll LJ, Rai AJ, et al. Three biomarkers identified from serum proteomic analysis for the detection of early stage ovarian cancer. ***Cancer Res*** 2004;64: 5882-90.
35. Wulfkuhle JD, McLean KC, Paweletz CP, Sgroi DC, Trock BJ, Steeg PS, et al. New approaches to proteomic analysis of breast cancer. ***Proteomics*** 2001;1:1205-15.
36. Baggerly KA, Morris JS, Coombes KR. Reproducibility of SELDI-TOF protein patterns in serum: comparing datasets from different experiments. ***Bioinformatics*** 2004;20:777-85.
37. Bons JAP, Wodzig WKWH, De Boer D, Drent M, Van Dieijen-Visser MP. Application of SELDI-TOF-MS in protein profiling: promises and Pitfalls. ***Jugoslov Med Biohem*** 2006;25: 201-10.