



Development and validation of a decision tree classification model for the essential hypertension based on serum protein biomarkers

Zongqiang Han¹, Lina Wen²

¹Department of Laboratory Medicine, Beijing Xiaotangshan Hospital, Beijing, China; ²Department of Clinical Nutrition, Beijing Shijitan Hospital, Capital Medical University, Beijing, China

Contributions: (I) Conception and design: Z Han; (II) Administrative support: Z Han; (III) Provision of study materials or patients: Z Han; (IV) Collection and assembly of data: Both authors; (V) Data analysis and interpretation: Both authors; (VI) Manuscript writing: Both authors; (VII) Final approval of manuscript: Both authors.

Correspondence to: Zongqiang Han. Department of Laboratory Medicine, Beijing Xiaotangshan Hospital, Beijing, China. Email: zongqianghan@sina.com.

Background: Essential hypertension (EH) is a key risk factor for cardiovascular disease. However, the etiology of EH is complex and unknown. So far, there is no good protein biomarker for screening EH. The purpose of this study was to discover potential biomarkers for EH by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) and establish a decision-tree classification model.

Methods: A total of 108 patients with clinically confirmed EH and 105 HC were enrolled in the present study from September 2020 to April 2021 and were randomly divided into the training group and the blind-test group. The serum protein expression profiles were performed using MALDI-TOF MS combined with magnetic beads with weak cation exchange (MB-WCX). The training group, which comprised 54 EH patients and 53 HC, was used to screen the statistically differential protein peaks by SPSS 19.0 and construct a decision-tree classification model by C5.0 algorithms of SPSS Modeler 18.0. All protein peak intensities of samples in the blind-test group, which comprised 54 EH patients and 52 HC, were used to verify the diagnostic capabilities of the model by classification model.

Results: EH patients had higher age, systolic and diastolic blood pressures than HC group. The intensities of 60 protein peaks differed significantly between the EH patients and HC. An optimal decision-tree classification model of EH was successfully established with mass-to-charge ratios of 1,326.7, 1,785.3, 4,228.0, and 8,963.8 as differential protein peaks by the software analysis. The decision-tree classification model was able to distinguish between EH patients and HC and had a sensitivity of 94.44%, a specificity of 94.33%, an accuracy of 94.39%, and an area under the receiver operating characteristic (ROC) curve of 0.96. The blind-test results indicated a sensitivity of 87.04%, a specificity of 88.46%, an accuracy of 87.74%, and an area under the ROC curve of 0.928.

Conclusions: MALDI-TOF MS combined with MB-WCX can be used to screen for serum differential protein expression profiles in EH patients. The decision-tree classification model based on mass-to-charge ratios of 1,326.7, 1,785.3, 4,228.0, and 8,963.8 could provide a new and reliable method for screening and identifying EH with high sensitivity and specificity.

Keywords: Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS); essential hypertension (EH); protein biomarkers

Submitted Jul 04, 2022. Accepted for publication Sep 02, 2022.

doi: 10.21037/atm-22-3901

View this article at: <https://dx.doi.org/10.21037/atm-22-3901>

Introduction

Hypertension is a common and frequently occurring chronic disease that seriously endangers human health in today's society. The latest epidemiological study showed that the prevalence of adult hypertension in China is 27.9% (standardized rate: 23.2%), and the prevalence continues to increase (1). In recent years, the awareness rate (51.6%), treatment rate (45.8%), and control rate (16.8%) of hypertension patients in China have improved, but the overall level is still relatively low. Hypertension is a risk factor for many diseases, including heart failure, left ventricular hypertrophy, stroke, renal failure, and retinopathy (1), and is a major medical problem worldwide. The target organ damage caused by hypertension is an important reason for the decrease in quality of life, disability, and even death of patients.

Essential hypertension (EH) refers to the high blood pressure of unknown cause and persistent elevated blood pressure is the main clinical manifestation, accounts for about 90–95% of all hypertension cases (2), and requires a diagnosis of exclusion. EH is a complex cardiovascular and cerebrovascular disease. Due to its complexity and heterogeneity, it is particularly difficult to study the pathogenesis of hypertension. Previous study has shown that genetics, age, gender, smoking, alcohol consumption, obesity, dietary habits, and region are all correlated with the incidence of EH (3).

Hypertension is usually diagnosed based on blood pressure as measured by a sphygmomanometer. Despite the simple method of blood pressure measurement, the blood pressure measurements are not always stable and elevated. However, pathological changes that occur in the body often precede an increase in blood pressure. Meanwhile, the onset of EH symptoms is delayed and the screening of EH is often ignored. Thus, the development of reliable diagnostic biomarkers for EH could help clinicians to intervene at an early stage and thus benefit patients. However, the pathogenesis of EH is still unclear, and there is a lack of early diagnosis methods. The discovery of the essential proteins involved in the occurrence and development of hypertension will not only lead to a deeper understanding of the pathogenesis of hypertension but will also provide new targets for drug exploration. Identifying early biomarkers of EH is also essential to prevent its progression and predict future cardiovascular events associated with EH in health and unhealthy individuals.

Some serum protein markers, such as leptin (4),

vascular cell adhesion molecule (5), intercellular adhesion molecule (6), high sensitivity C-reactive protein (7), tumor necrosis factor α (8), have been proved to be associated with EH. However, as a single protein marker, it is difficult to correctly classify normotensive and EH patients in the clinic due to lack of sensitivity and specificity. However, prior studies have rarely attempted to screen the specific serum protein biomarkers for EH by proteomics methods based on mass spectrometry technologies. Among these methods, matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) combined with magnetic beads is an advanced technology for directly analyzing complex clinical sample and screening disease protein biomarkers. It can detect low molecular weight proteins that may contain important biological significance, as they have very low content in serum and are difficult to measure them by traditional protein detection technology. A panel of low molecular weight proteins used for constructing disease classification models is considered as potential biomarkers. It can also improve the specificity and sensitivity of early diagnosis of disease.

MALDI-TOF MS, which is an important proteomics technology, is mainly composed of the following 3 parts: a matrix-assisted laser desorption/ionization ion source, a time-of-flight mass analyzer, and a detector. The basic principle is to mix the measured sample with a solution of a saturated small molecule matrix compound to make the measured sample evenly dispersed in the matrix. The sample is then placed on the target plate, naturally crystallized and dried, and put in the ion source chamber. In the high vacuum state (about 10^{-7} mbar), the pulsed laser is irradiated on the target sample. The matrix absorbs energy from the laser and converts it to an ionized state and transmitted to the sample molecules. Charge transfer occurs between the matrix and the sample to ionize the sample molecules. Next, under the action of the electric field, the sample is accelerated into the time-of-flight mass analyzer, and the mass-charge ratio (m/z) is calculated from the time of flight to the detector to conduct a qualitative or quantitative analysis of the sample.

Magnetic beads (also referred to as nanomagnetic beads) are usually composed of a magnetic core and a polymer shell on the outside of the core. Magnetic beads can be separated from the surrounding medium in the magnetic field. They can bind to a variety of biologically active substances, such as nucleic acids, enzymes, receptors, antigens, and antibodies, by modification to the microspheres surface. There are several types of magnetic beads, such as the weak

Table 1 Basic characteristics of the subjects

Parameters	EH	HC	P value
Sample, n	108	105	–
Male/female	79/29	36/69	<0.001
Age range, years	24–80	22–62	<0.001
Mean age, years (mean \pm SD)	53.5 \pm 11.7	34.7 \pm 7.7	<0.001
Mean SBP, mmHg (mean \pm SD)	145.0 \pm 16.5	110.6 \pm 6.4	<0.001
Mean DBP, mmHg (mean \pm SD)	86.7 \pm 12.9	69.0 \pm 5.3	<0.001
Training set	54	53	–
Blind-test set	54	52	–

EH, essential hypertension; HC, healthy controls; SD, standard deviation; SBP, systolic blood pressure; DBP, diastolic blood pressure.

cation exchange (WCX) type, weak anion exchange (WAX) type, hydrophobic interaction of carbon binding group type (e.g., C3, C8, and C18), immobilized metal ion affinity chromatograph with Cu(2+) or Fe(3+) (IMAC-Cu, Fe) type, and immunoaffinity chromatography on immobilized protein G (IAC Prot G) type (9). Bioactive substances can be fixed to magnetic beads, which have large surface areas and can capture small peptides and proteins (10,11).

MALDI-TOF MS technology combined with magnetic beads has been widely used in the study of a variety of diseases and has led to the discovery of a variety of protein biomarkers, including for cancer (12), immune diseases (13), and endocrine diseases (14). MALDI-TOF MS technology combined with magnetic beads can directly detect all kinds of complex biological samples (e.g., urine, blood, cerebrospinal fluid, tissue biopsies, and even tear samples). Additionally, the required sample amount is very small; only 0.5 to 5 μ L is required per analysis. The range of detectable molecular weight is wide, especially for proteins with molecular weights <20 kDa. Thus, it can also be used to detect low abundance proteins or peptides. Additionally, MALDI-TOF MS technology also has the advantages of a fast analysis speed, high sensitivity, ease of use, cheap consumables, and wide coverage. Thus, the platform may be suitable for screening a large number of samples. Finally, one or a group of biomarkers from differential protein peaks related to the disease can be screened by a software analysis, which provides the best combination of laboratory indicators for disease diagnosis. However, the main drawback is that MALDI-TOF MS cannot detect the amino acid sequence of related proteins/peptides (15).

Currently, there is no effective serum protein biomarker for the early diagnosis of hypertension. In our study, MALDI-TOF MS combined with magnetic beads with weak cation exchange (MB-WCX) was used to detect the serum proteomic profiles of the 108 EH patients and 105 healthy controls (HC). The samples were randomly divided into the training group and blind-test group. The training group was used to screen differential protein peaks in EH and establish a decision-tree classification model. The blind-test group was applied to verify the sensitivity, specificity, and accuracy of the decision-tree classification model. We present the following article in accordance with the STARD reporting checklist (available at <https://atm.amegroups.com/article/view/10.21037/atm-22-3901/rc>).

Methods

Patients and controls characteristics

This is a case-control study. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). The study was approved by the Beijing Xiaotangshan Hospital Ethics Committee (No. 2019-25) and the individual consent for this retrospective analysis was waived. The sample size was set accordingly (16,17). A total of 108 EH patients and 105 HC were included in this study. After randomization, the training group comprised 54 EH patients and 53 HC, and the blind-test group comprised 54 EH patients and 52 HC. The collection time was from September 2020 to April 2021 for the case group (see *Table 1*). To be eligible for inclusion in the EH group, the subject had to meet the following inclusion criteria: (I) have been diagnosed with EH by specialists according to

the diagnostic criteria for the 2018 Chinese Guidelines for Prevention and Treatment of Hypertension [i.e., following 3 measurements of blood pressure in the clinic on different days, the patient had systolic blood pressure ≥ 140 mmHg, and/or diastolic blood pressure ≥ 90 mmHg (1 mmHg = 0.133 kPa)]; (II) be aged 18–80 years; (III) have no damage to the target organs, such as the heart, liver, brain, kidney, and lungs; and (IV) despite taking antihypertensive drugs, have a blood pressure level that still met the diagnostic criteria after stopping the drugs for 2 weeks. Subjects were excluded from the EH group if they met any of the following exclusion criteria: (I) had secondary hypertension; (II) had a serious primary disease, such as a cardiovascular, cerebrovascular, liver, kidney, digestive, endocrine, and hematopoietic system disease; (III) had epilepsy combined with obstructive sleep apnea hypopnea syndrome; (IV) was a psychiatric patient with high anxiety and depression; (V) was a pregnant or a lactating woman, or a woman of childbearing age who did not take contraceptive measures; (VI) was at risk of infection (e.g., HIV carriers, hepatitis B surface antigen positive, or syphilis positive patients); and/or (VII) had a progressive malignant tumor or other serious wasting disease that could easily be complicated by infection and bleeding; and/or (VIII) had undergone a major operation, such as hysterectomy, thyroidectomy, or ventricular septal defect. The HC group comprised individuals who had health examinations at our hospital during the corresponding period. To be eligible for inclusion in the HC group, the subjects had to meet the following inclusion criteria: (I) be aged 18–80 years; and/or (II) be healthy with no previous organic or functional diseases, and no recent colds or other abnormalities. Subjects were excluded from the HC group if they met any of the following exclusion criteria: (I) had a known family history of hypertension, type 1 or type 2 diabetes mellitus, cerebrovascular diseases; (II) were a pregnant or lactating woman; and/or (III) had a psychiatric history.

Instruments and reagents

MALDI-TOF MS (ProteinChip Biology System II-*c*, PBS II-*c*) was purchased from Ciphergen Biosystems company (Austin, TX, USA), and Au chip was purchased from BioRad Laboratories, Inc. (Hercules, CA, USA). MB-WCX were purchased from Bruker Daltonics company (USA). Sodium acetate, urea, dithiothreitol (DTT), acetonitrile (ACN), Tris-HCl (pH9.0), Trifluoroacetic acid (TFA), sinapinic acid (SPA), water (HPLC grade), 3-[(3-choleamidopropyl)

dimethylamino]-1-propanesulfonate (CHAPS), and N-2-hydroxyethylpiperazine-N-2-ethanesulfonic acid (HEPES) were purchased from Sigma Aldrich (USA).

Serum sample preparation

All the blood samples were collected in 3-mL serum vacuum tubes (Becton Dickinson Vacutainer Systems, USA) under early morning fasting blood and allowed to clot for 1 h at 4 °C, and then centrifuged at 4,000 rpm for 20 min and stored at –80 °C for further analysis.

Sample pretreatment

Frozen serum samples, which were taken out of the refrigerator at –80 °C, were defrosted on ice and centrifuged at 10,000 rpm at 4 °C for 5 min. Next, 10 μ L of serum and 20 μ L of U9 buffer (9 M Urea, 0.2% CHAPS, 0.1% DTT, and 50 M Tris-HCl, pH 9.0) were added to the labeled 1.5-mL centrifuge tube, and thoroughly mixed. After shaking in an ice bath for 30 min (500 rpm), 370 μ L of sodium acetate buffer (100 mM, pH4.5) was added and mixed immediately.

Magnetic bead pretreatment, sample loading and elution

A total of 50 μ L MB-WCX (50 mg/mL) was added to the 200- μ L polymerase chain reaction (PCR) tube that was placed in the magnetic bead separator for 1 min, and the liquid was then removed. Next, 100 μ L of sodium acetate buffer (100 mM, pH4.5) was added to the PCR tube, which was taken off the magnetic bead separator and carefully mixed with the magnetic beads and left standing for 5 min. The PCR tube was placed in the magnetic bead separator, and the liquid was removed, and the procedure was repeated once. Next, 100 μ L of pretreated sample was added to the activated magnetic beads, and shaken at 500 rpm for 1 min. The PCR tube was placed in the magnetic bead separator for 2 min and the unbound sample was removed. Then, 100 μ L of sodium acetate buffer (100 mM, pH4.5) was added to the PCR tube, which was taken from magnetic bead separator and carefully mixed with the magnetic beads and left standing for 5 min. The PCR tube was placed in the magnetic bead separator and the liquid was removed, and the procedure was repeated once. Next 10 μ L of elution (0.5% TFA) was added to the PCR tube for 15 min, and it was then placed on the magnetic bead separator.

Table 2 Evaluation of the precision of the mass spectrometer by the CVs of selected protein peaks

Protein peak, m/z	CV (%)
2,712.3	7.2
4,070.5	1.3
4,576.7	14.9
5,321.4	15.4
5,620.7	12.4
5,892.4	12.5
6,624.3	11.6
7,772.2	12.1
8,576.7	13.5
8,703.5	8.0
9,210.2	7.9

CV, coefficient of variation.

MALDI-TOF MS analysis

Eluent (5 μ L) was pipetted into another PCR tube and thoroughly mixed with 5 μ L of SPA (50% ACN, 0.5% TFA). Next, 1 μ L of the mixture was pipetted onto the Au chip (Bio Rad, USA) and left to dry naturally at room temperature. The prepared Au chip was immediately tested by MALDI-TOF MS in the positive-ion mode. The mass spectrometry parameter settings were as follows: laser intensity: 180; detection sensitivity: 8; optimized range of the relative molecular mass: 2,000–10,000 Da; and maximum relative molecular mass: 50,000 Da. Each point on the chip was collected 80 times. Instrument external calibration was performed by standard procedures using the all-in-one peptide molecular mass standards before the MS analysis, which contained 5 polypeptides [i.e., arginine 8-vasopressin (1,084.247 Da), somatostatin (1,637.903 Da), bovine insulin B chain (3,495.941 Da), human insulin (5,807.653 Da), and hirudin (7,033.614 Da)], with the protein molecular weight error <0.1%. To assess the stability and repeatability of the MALDI-TOF MS, 8 within-run analyses were performed using quality control samples (collected from 10 EH patients) during the experiment, to ensure a coefficient of variation of mean peak intensity of quality control protein <16%.

Data processing and statistical analysis

The total ion flow normalization method in Ciphergen ProteinChip 3.2.1 software was used to normalize the obtained

original mass spectrograms with mass-to-charge ratios ranging from 1,000 to 50,000 Da. Biomarker wizard software 3.1 was used to label the protein peaks. IBM SPSS Statistics for Windows, Version 19.0 software was used to statistically analyze the original data of protein peak intensity. Categorical variables were compared in the two groups using the χ^2 test. All the continuous variables are presented as the mean \pm standard deviation. An independent samples *t*-test (two-sided) or Mann-Whitney U test (based on the normality of data distribution) was used for comparisons between the groups. P values <0.05 were considered statistically significant. C5.0 algorithms of IBM SPSS Modeler 18.0 was used to construct the EH decision-tree classification model, and the receiver operating characteristic (ROC) curve and area under ROC curve (AUC) values. For C5.0 algorithms, expert mode was selected and the parameter settings were as follows: pruning severity: 75; minimum number of records per child branch: 2; use global pruning. The sensitivity and specificity of the model were obtained. The sensitivity was regarded as the percentage of EH correctly predicted, and the specificity was regarded as the percentage of HC correctly predicted. Accuracy was defined as the ability of the model to correctly classify EH and HC. ROC curve and AUC values were calculated to assess the ability of the model to identify EH.

Results

Demographic and clinical characteristics of study subjects

EH patients had higher age, systolic and diastolic blood pressures than HC group (P<0.001). The detailed demographic and clinical characteristics of all study subjects are presented (see *Table 1*).

Mass spectrometer precision evaluation

The mass spectrometer precision was evaluated using a quality control serum sample collected from 10 EH patients. The serum sample was repeatedly detected 8 times within-run assays by MALDI-TOF MS. The coefficients of variation (CVs) of the 11 protein peaks relative intensity ranged from 1.3% to 15.4% (see *Table 2*). The CV range was considered acceptable for measuring the serum sample by MALDI-TOF MS (PBS II-c).

Detection of serum protein expression profile

MALDI-TOF MS was used to detect 213 serum samples

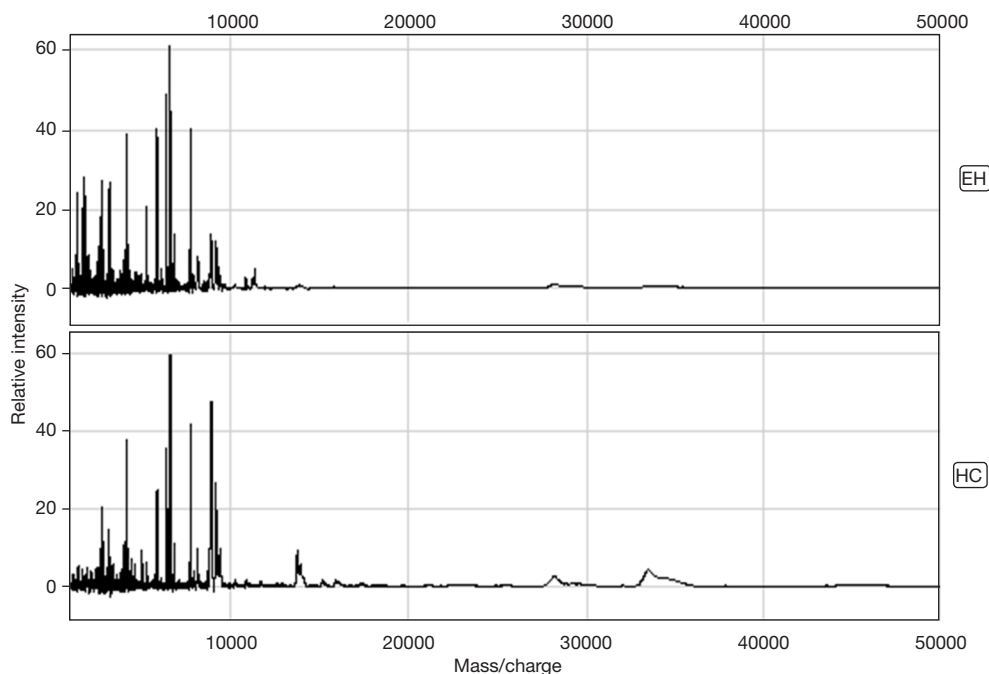


Figure 1 Representative mass spectrogram of single serum samples from EH patients and HC, respectively, detected by MALDI-TOF MS combined with MB-WCX, showing the protein m/z between 1,000 and 50,000. EH, essential hypertension; HC, healthy controls; MALDI-TOF MS, matrix-assisted laser desorption/ionization time-of-flight mass spectrometry; MB-WCX, magnetic beads with weak cation exchange.

from 108 EH patients and 105 HC. The original protein profiles were standardized and analyzed by Biomarker Wizard Software 3.1.0. A total of 106 protein peaks were detected from 1,000 to 50,000 Da (see *Figure 1*). This showed the effectiveness of the MALDI technique in isolating and detecting low molecular weight proteins (<20,000 Da).

Serum protein profile analysis between EH group and HC group

The serum protein profiles of 54 EH patients and 53 HC in the training group were detected by MALDI-TOF MS. In total, 60 protein peaks differed significantly between the EH group and HC group ($P < 0.05$), among which 54 differential protein peaks had high expression and 6 differential protein peaks had low expression in the EH group (see *Table 3*).

Construction of the decision-tree classification model between EH group and HC group

The decision-tree classification model was constructed by

IBM SPSS Modeler 18.0 software. The data used in the SPSS Modeler 18.0 software was from Biomarker Wizard Software 3.1.0. In total, 4 differential protein peaks (i.e., 1,326.7, 1,785.3, 4,228.0, and 8,963.8 m/z) were selected to construct the optimal decision-tree classification model for the diagnosis of EH (see *Figure 2*). Compared to the HC group, the expression of the differential protein peaks of 1,326.7, 1,785.3, 4,228.0 and 8,963.8 m/z were all up-regulated in the EH group (see *Figure 3*). The 107 serum samples were divided into 5 terminal nodes. The decision-tree classification model based on these 4 biomarkers produced a good classification between the EH group and the HC group, with a sensitivity of 94.44%, specificity of 94.33%, and accuracy of 94.39% (see *Table 4*). The AUC of the decision-tree classification model was 0.96 (see *Figure 4A*).

Diagnostic performance of decision-tree classification model

The decision-tree classification model was able to distinguish between the EH group and the HC group. The accuracy of the optimal decision-tree classification model

Table 3 The mean expression level of differential protein peaks in the EH group and HC group

Protein peak, m/z	z/t	P value	EH group (n=54)		HC group (n=53)	
			Mean	SD	Mean	SD
1,174.7	-4.617	<0.001	1.06	1.96	2.60	2.17
1,326.7	-5.904	<0.001	8.01	5.88	3.02	1.95
1,442.3	-5.028	<0.001	1.11	2.71	3.02	2.69
1,456.2	-6.299	<0.001	5.40	5.64	0.53	0.64
1,698.7	-5.960	<0.001	4.22	4.41	0.61	0.62
1,785.3	-6.805	<0.001	6.90	6.30	1.01	0.91
2,030.7	-4.987	<0.001	3.57	3.33	1.20	1.04
2,079.4	-2.050	0.040	1.56	2.07	0.74	1.04
2,286.4	2.190	0.031	0.96	1.05	1.40	1.06
2,512.3	-3.147	0.002	0.90	1.25	1.62	1.45
2,552.2	-4.623	<0.001	4.12	3.68	1.26	1.12
2,652.2	-5.265	<0.001	4.92	4.01	1.30	1.88
2,696.8	-3.967	<0.001	2.99	3.36	1.08	1.10
2,767.4	-4.654	<0.001	9.82	7.19	3.99	4.54
2,881.0	-4.710	<0.001	15.40	10.36	6.49	7.02
2,948.5	-4.742	<0.001	5.43	4.26	1.81	2.73
3,170.7	-4.623	<0.001	12.80	9.44	4.74	4.73
3,206.5	-4.162	<0.001	8.21	8.59	1.38	1.76
3,275.8	-5.564	<0.001	11.46	11.05	1.18	2.05
3,331.0	-3.087	0.003	4.47	4.47	2.32	2.51
3,386.0	-4.587	<0.001	5.93	4.76	2.47	2.80
3,458.6	-4.885	<0.001	5.65	4.26	2.39	2.65
3,501.9	-4.168	<0.001	2.37	2.64	0.93	1.54
3,893.6	-2.455	0.014	0.99	1.76	0.47	0.84
4,018.6	-3.209	0.001	-0.20	0.54	0.40	1.45
4,073.3	-3.689	<0.001	2.84	2.51	1.59	2.55
4,109.5	-4.530	<0.001	4.56	3.28	2.10	3.59
4,228.0	-4.567	<0.001	20.44	13.90	8.02	10.25
4,293.3	-3.022	0.003	3.74	5.36	1.64	2.85
4,486.8	-3.832	<0.001	2.37	3.05	1.13	1.72
4,662.4	-3.097	0.002	2.87	3.01	1.38	1.58
4,804.5	-5.570	<0.001	2.09	2.28	0.13	0.42
4,838.2	-4.474	<0.001	1.08	1.91	0.52	2.36
4,980.2	-5.408	<0.001	2.08	2.50	0.32	1.23

Table 3 (continued)

Table 3 (continued)

Protein peak, m/z	z/t	P value	EH group (n=54)		HC group (n=53)	
			Mean	SD	Mean	SD
5,086.2	-4.094	<0.001	3.16	2.09	1.62	1.90
5,358.0	-4.661	<0.001	13.28	13.07	2.03	2.48
5,407.0	-3.196	<0.001	0.71	0.77	0.36	0.64
5,928.3	-4.748	<0.001	26.19	19.72	6.54	7.35
6,135.1	-5.718	<0.001	3.05	3.12	0.56	0.71
6,460.9	-4.486	<0.001	29.68	17.59	15.37	13.78
6,658.5	-4.891	<0.001	45.66	17.87	26.85	18.87
6,867.9	-4.760	<0.001	8.10	5.16	3.82	3.51
7,190.5	-4.318	<0.001	2.12	1.37	1.78	4.06
7,797.2	-3.913	<0.001	22.75	16.54	9.34	9.82
7,991.3	-4.878	<0.001	1.91	1.52	0.77	0.77
8,172.0	-4.143	<0.001	5.33	4.00	2.35	2.37
8,248.7	-4.156	<0.001	3.43	3.28	1.93	2.76
8,616.8	-4.274	<0.001	0.74	0.63	0.34	0.41
8,837.5	-3.536	0.001	4.19	2.72	2.65	1.65
8,963.8	-3.309	0.001	16.97	12.18	11.37	11.41
9,322.6	-3.570	<0.001	9.12	8.76	3.15	3.26
10,297.9	-2.978	0.003	1.10	1.24	0.46	0.43
10,918.5	-2.960	0.003	0.84	0.93	0.53	0.78
11,104.1	-2.480	0.013	0.45	0.44	0.34	0.42
22,997.0	-1.994	0.046	0.24	0.25	0.25	0.15
30,392.9	-4.368	<0.001	0.20	0.17	0.07	0.06
33,903.8	-2.014	0.047	1.02	0.67	0.76	0.66
34,168.0	-3.589	<0.001	0.94	0.60	0.63	0.56
34,502.2	-3.284	0.001	0.91	0.56	0.64	0.53
45,158.2	-2.842	0.005	0.18	0.11	0.12	0.10

EH, essential hypertension; HC, healthy controls; SD, standard deviation.

was determined by comparing the clinical diagnosis to the model's judgment of each sample. In the training set, 51 of the 54 EH patients and 50 of the 53 HC were accurately identified by the decision-tree classification model. The blind-test results showed that the sensitivity, specificity, and accuracy of the model for distinguishing between the EH group and HC group were 87.04%, 88.46%, and 87.74%, respectively (see *Table 4*). The AUC of the decision-tree

classification model was 0.928 (see *Figure 4B*).

Discussion

Globally, 874 million adults have unsatisfactory systolic blood pressure levels ≥ 140 mmHg, and approximately 1/4 of adults suffer from hypertension (18). The Global Burden of Disease study showed that sub-optimal blood pressure (i.e.,

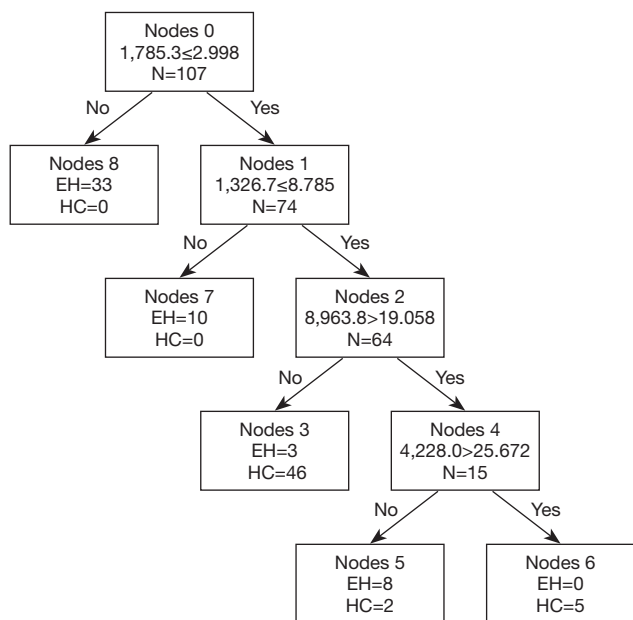


Figure 2 Diagram of the decision-tree classification model for EH patients and HC. EH, essential hypertension; HC, healthy controls.

>110–115 mmHg) remains the single largest risk factor for the global burden of disease and global mortality, resulting in 9.4 million deaths and 212 million losses of healthy lives each year (8.5% of the global total) (19). Hypertension is a genetic disease with a complex pathogenesis, it has involved many factors, a high incidence, causes great harm, and has a strong genetic correlation. The etiology of 90% of patients with hypertension is unknown, which is called EH (20).

The pathogenesis of EH is very complex, and the cause of EH has not been completely expounded. A number of factors are thought to be involved in the pathogenesis of EH. First, genetic genes. Human EH is a multi-gene inherited disease. More than 150 genes are related to hypertension; for example, the rs752107 (C>T) polymorphism of the Wnt family member 3A (*WNT3A*) gene increases the risk of EH (21), while the rs671 polymorphism of the aldehyde dehydrogenase 2 family member (*ALDH2*) gene reduces the risk of EH (22). Second, increased sympathetic nerve activity. Sympathetic nerve fibers are extensively spread throughout the cardiovascular system. Catecholamines released by increased sympathetic nerve excitability mainly

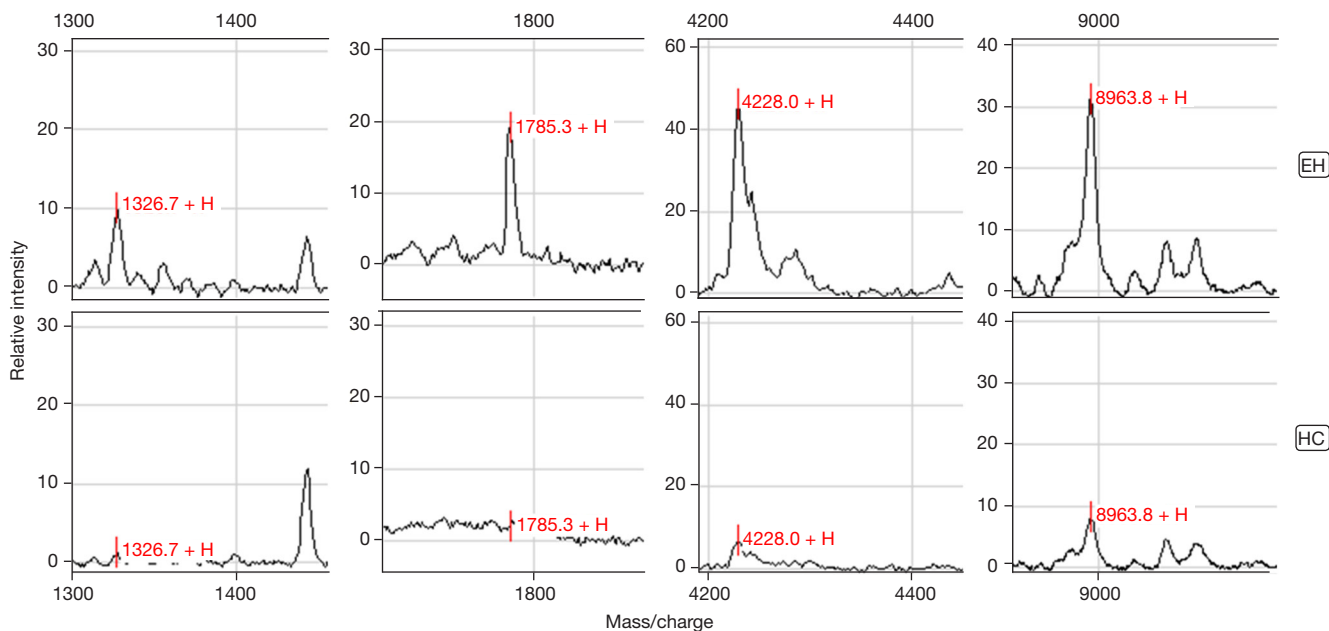


Figure 3 The 4 selected differential expression protein peaks of 1,326.7, 1,785.3, 4,228.0, and 8,963.8 m/z in the EH and HC groups. EH, essential hypertension; HC, healthy controls.

Table 4 The model diagnostic performance verification for EH by the blind test

Group/clinical group	Cases, n	Correct cases	Correct rate	Sensitivity	Specificity
Training set					
EH	54	51	–	–	–
HC	53	50	–	–	–
Blind-test set					
EH	54	47	–	–	–
HC	52	46	–	–	–

EH, essential hypertension; HC, healthy controls.

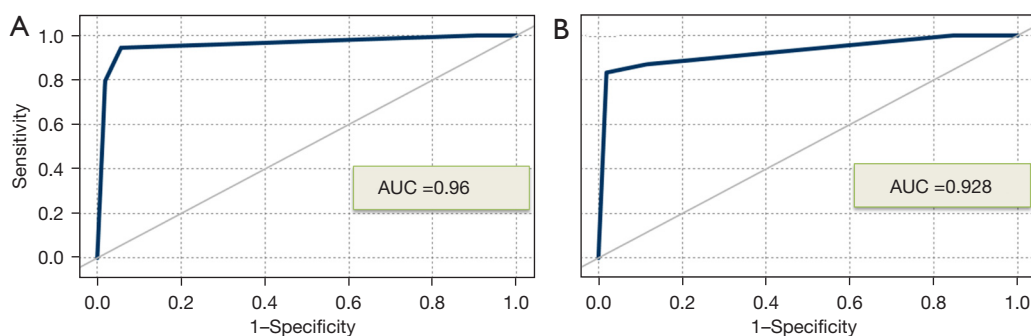


Figure 4 Evaluation of the diagnostic ability of the decision-tree classification model of the EH group and HC group by ROC curves and AUC values. (A) For the training set comprising 54 EH patients and 53 HC, the AUC value was 0.96. (B) For the blind-test set comprising 54 EH patients and 52 HC, the AUC value was 0.928. AUC, area under curve; EH, essential hypertension; HC, healthy controls; ROC, receiver operating characteristic.

act on the heart, resulting in an accelerated heart rate, stronger myocardial contractility, and increased cardiac output. The use of catecholamine on alpha-adrenergic receptors in the vascular wall constricts arterioles, increases peripheral resistance, and increases blood pressure (23). Third, renal function. The main physiological function of the kidney is to regulate water, electrolytes, blood volume, and excrete metabolites in the body. Renal dysfunction can lead to water and sodium retention and increased blood volume, resulting in increased blood pressure (24). Fourth, the vascular mechanism (i.e., endothelial dysfunction and the nitric oxide pathway). Impaired endothelial function and nitric oxide can lead to elevated blood pressure (25,26). Fifth, the renin-angiotensin-aldosterone system (RAAS). The enhanced function of RAAS promotes vasoconstriction, water, and sodium retention, and myocardial and vascular remodeling (27). Sixth, obesity. Weight gain increases sympathetic activity to burn fat, but this sympathetic overactivity leads to hypertension (28).

Seventh, hyperinsulinemia and hyperuricemia. Insulin can increase blood pressure through a number of mechanisms (i.e., it can increase renal sodium reabsorption, activate the sympathetic nervous system, change transmembrane ion transport, and increase peripheral vascular resistance) (29). Hyperuricemia can induce vascular insulin resistance and endothelial dysfunction, leading to hypertension (30).

Proteomic studies on EH have been published; however, the number is limited. In the Anglo-Scandinavian Heart Outcome Trial, urinary proteomics was found to provide information about cardiovascular endpoints and predict coronary events in patients with hypertension (31). Molecular markers based on signaling molecules, growth factors, and angiogenic factors were identified by plasma proteomics studies in patients with hypertension and diabetes mellitus. These molecular markers can be used to predict the risk of renal disease in patients with hypertension or type 2 diabetes (32). The characteristics identified by the comprehensive proteomic analysis of 123

hypertensive patients with chronic RAAS inhibition could predict the progression of proteinuria. These features include the proteins involved in the immune inflammatory response and endoplasmic reticulum stress activation. Further, studies have found a link between inflammation and proteinuria development in controlled-hypertensive patients by blocking RAAS (33). In a serum protein analysis of 118 hypertensive patients, comprising 9 diabetes mellitus patients, 9 coronary disease patients, and 6 myocardial infarction patients, Gajjala *et al.* found 27 differential protein peaks and identified 18 of them (17). The present study was limited to EH patients, the interference of other diseases was strictly excluded, and MB-WCX combined with MALDI-TOF MS technology was applied to detect the serum protein to ensure accurate and reliable results.

By analyzing the serum proteomic profiles of EH patients and HC in the training group, 60 differential protein peaks were screened out. The IBM SPSS Modeler 18.0 machine-learning algorithm was used to analyze the training set data and establish a decision-tree classification model for EH involving 4 potential biomarkers (34). The accuracy of the decision-tree classification model for EH was 94.39%. The ROC curves proved that the established diagnostic model could effectively distinguish between EH patients and HC. The validation data of the blind-test set showed that the decision-tree could identify 47 of the 54 cases of EH and 46 of the 52 cases of HC. The high sensitivity and specificity of the established decision-tree classification model is promising.

We screened potential biomarkers related to EH and established a decision-tree classification model. However, the limitations of our study relate to the number of samples and the mass spectrometry analysis methods. MALDI-TOF MS technology is an important tool for proteomics research, and can obtain rich spectral information. However, the selected biomarkers cannot be identified. Additionally, this study was a single center study. A large number of clinical samples are still needed to verify the applicability of the model established in this study.

In conclusion, the low molecular weight serum protein profiles associated with EH were examined using MB-WCX combined with MALDI-TOF MS technology, and the differential protein peaks were screened out. By selecting 4 biomarkers, a differential diagnosis model was established for EH patients and HC, which had high sensitivity and high specificity for EH. Our results provide novel insights into changes in serum protein profiles during the development of EH. Potential biomarkers may be involved in the pathologic process of EH.

Acknowledgments

Funding: This work was supported by the Beijing Municipal Administration of Hospitals Incubating Program (No. PX2020078) and the Beijing Xiaotangshan Hospital Scientific Research Project (No. Tang2021-01).

Footnote

Reporting Checklist: The authors have completed the STARD reporting checklist. Available at <https://atm.amegroups.com/article/view/10.21037/atm-22-3901/rc>

Data Sharing Statement: Available at <https://atm.amegroups.com/article/view/10.21037/atm-22-3901/dss>

Conflicts of Interest: Both authors have completed the ICMJE uniform disclosure form (available at <https://atm.amegroups.com/article/view/10.21037/atm-22-3901/coif>). The authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). The study was approved by the Beijing Xiaotangshan Hospital Ethics Committee (No. 2019-25) and individual consent for this retrospective analysis was waived.

Open Access Statement: This is an Open Access article distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with the strict proviso that no changes or edits are made and the original work is properly cited (including links to both the formal publication through the relevant DOI and the license). See: <https://creativecommons.org/licenses/by-nc-nd/4.0/>.

References

1. Joint Committee for Guideline Revision. 2018 Chinese Guidelines for Prevention and Treatment of Hypertension-A report of the Revision Committee of Chinese Guidelines for Prevention and Treatment of Hypertension. *J Geriatr Cardiol* 2019;16:182-241.

2. Oparil S, Acelajado MC, Bakris GL, et al. Hypertension. *Nat Rev Dis Primers* 2018;4:18014.
3. Li AL, Peng Q, Shao YQ, et al. The interaction on hypertension between family history and diabetes and other risk factors. *Sci Rep* 2021;11:4716.
4. Bielecka-Dabrowa A, Bartłomiejczyk MA, Sakowicz A, et al. The Role of Adipokines in the Development of Arterial Stiffness and Hypertension. *Angiology* 2020;71:754-61.
5. Shalia KK, Mashru MR, Vasvani JB, et al. Circulating levels of cell adhesion molecules in hypertension. *Indian J Clin Biochem* 2009;24:388-97.
6. Biel V, Novak J, Rimalova V, et al. Levels of endothelial substances in patients with newly identified hypertension compared with healthy controls. *Biomed Pap Med Fac Univ Palacky Olomouc Czech Repub* 2021;165:395-401.
7. Bisaria S, Terrigno V, Hunter K, et al. Association of Elevated Levels of Inflammatory Marker High-Sensitivity C-Reactive Protein and Hypertension. *J Prim Care Community Health* 2020;11:2150132720984426.
8. Banaszak B, Świętochowska E, Banaszak P, et al. Endothelin-1 (ET-1), N-terminal fragment of pro-atrial natriuretic peptide (NTpro-ANP), and tumour necrosis factor alpha (TNF- α) in children with primary hypertension and hypertension of renal origin. *Endokrynol Pol* 2019;70:37-42.
9. Chai Z, Bi H. Capture and identification of bacteria from fish muscle based on immunomagnetic beads and MALDI-TOF MS. *Food Chem X* 2022;13:100225.
10. Zhu N, Xing X, Cao L, et al. Study on the Diagnosis of Gastric Cancer by Magnetic Beads Extraction and Mass Spectrometry. *Biomed Res Int* 2020;2020:2743060.
11. Hu Z, Tian Y, Li J, et al. Urinary Peptides Associated Closely with Gestational Diabetes Mellitus. *Dis Markers* 2020;2020:8880034.
12. Ding D, Chen M, Xiao X, et al. Novel serum peptide model revealed by MALDI-TOF-MS and its diagnostic value in early bladder cancer. *Int J Biol Markers* 2020;35:59-66.
13. Ma D, Liang N, Zhang L. Establishing Classification Tree Models in Rheumatoid Arthritis Using Combination of Matrix-Assisted Laser Desorption/Ionization Time-of-Flight Mass Spectrometry and Magnetic Beads. *Front Med (Lausanne)* 2021;8:609773.
14. Hu Z, Hou J, Zhang M. Levels of inter-alpha-trypsin inhibitor heavy chain H4 urinary polypeptide in gestational diabetes mellitus. *Syst Biol Reprod Med* 2021;67:428-37.
15. Zambonin C, Aresta A. MALDI-TOF/MS Analysis of Non-Invasive Human Urine and Saliva Samples for the Identification of New Cancer Biomarkers. *Molecules* 2022;27:1925.
16. Ameta K, Gupta A, Kumar S, et al. Essential hypertension: A filtered serum based metabolomics study. *Sci Rep* 2017;7:2153.
17. Gajjala PR, Jankowski V, Heinze G, et al. Proteomic-Biostatistic Integrated Approach for Finding the Underlying Molecular Determinants of Hypertension in Human Plasma. *Hypertension* 2017;70:412-9.
18. Forouzanfar MH, Liu P, Roth GA, et al. Global Burden of Hypertension and Systolic Blood Pressure of at Least 110 to 115 mm Hg, 1990-2015. *JAMA* 2017;317:165-82.
19. GBD 2015 Risk Factors Collaborators. Global, regional, and national comparative risk assessment of 79 behavioural, environmental and occupational, and metabolic risks or clusters of risks, 1990-2015: a systematic analysis for the Global Burden of Disease Study 2015. *Lancet* 2016;388:1659-724. Erratum in: *Lancet* 2017;389:e1.
20. Rivas AM, Pena C, Kopel J, et al. Hypertension and Hyperthyroidism: Association and Pathogenesis. *Am J Med Sci* 2021;361:3-7.
21. Ren H, Luo JQ, Ouyang F, et al. WNT3A rs752107(C > T) Polymorphism Is Associated With an Increased Risk of Essential Hypertension and Related Cardiovascular Diseases. *Front Cardiovasc Med* 2021;8:675222.
22. Mei XF, Hu SD, Liu PF, et al. ALDH2 Gene rs671 Polymorphism May Decrease the Risk of Essential Hypertension. *Int Heart J* 2020;61:562-70.
23. Tomoda F, Nitta A, Sugimori H, et al. Plasma and Urinary Levels of Nerve Growth Factor Are Elevated in Primary Hypertension. *Int J Hypertens* 2022;2022:3003269.
24. Chu Y, Zhou Y, Lu S, et al. Pathogenesis of Higher Blood Pressure and Worse Renal Function in Salt-Sensitive Hypertension. *Kidney Blood Press Res* 2021;46:236-44.
25. Martins AC, Santos AAD, Lopes ACBA, et al. Endothelial Dysfunction Induced by Cadmium and Mercury and its Relationship to Hypertension. *Curr Hypertens Rev* 2021;17:14-26.
26. Liu D, Yi L, Sheng M, et al. The Efficacy of Tai Chi and Qigong Exercises on Blood Pressure and Blood Levels of Nitric Oxide and Endothelin-1 in Patients with Essential Hypertension: A Systematic Review and Meta-Analysis of Randomized Controlled Trials. *Evid Based Complement Alternat Med* 2020;2020:3267971.
27. Chu HT, Li L, Jia M, et al. Correlation between serum microRNA-136 levels and RAAS biochemical markers in patients with essential hypertension. *Eur Rev Med Pharmacol Sci* 2020;24:11761-7.

28. Zhao W, Mo L, Pang Y. Hypertension in adolescents: The role of obesity and family history. *J Clin Hypertens (Greenwich)* 2021;23:2065-70.
29. Quesada O, Claggett B, Rodriguez F, et al. Associations of Insulin Resistance With Systolic and Diastolic Blood Pressure: A Study From the HCHS/SOL. *Hypertension* 2021;78:716-25.
30. Miyabayashi I, Mori S, Satoh A, et al. Uric Acid and Prevalence of Hypertension in a General Population of Japanese: ISSA-CKD Study. *J Clin Med Res* 2020;12:431-5.
31. Brown CE, McCarthy NS, Hughes AD, et al. Urinary proteomic biomarkers to predict cardiovascular events. *Proteomics Clin Appl* 2015;9:610-7.
32. Pena MJ, Jankowski J, Heinze G, et al. Plasma proteomics classifiers improve risk prediction for renal disease in patients with hypertension or type 2 diabetes. *J Hypertens* 2015;33:2123-32.
33. Baldan-Martin M, Mourino-Alvarez L, Gonzalez-Calero L, et al. Plasma Molecular Signatures in Hypertensive Patients With Renin-Angiotensin System Suppression: New Predictors of Renal Damage and De Novo Albuminuria Indicators. *Hypertension* 2016;68:157-66.
34. Zhang J, Tang W. Building a prediction model for iron deficiency anemia among infants in Shanghai, China. *Food Sci Nutr* 2020;8:265-72.

(English Language Editor: L. Huleatt)

Cite this article as: Han Z, Wen L. Development and validation of a decision tree classification model for the essential hypertension based on serum protein biomarkers. *Ann Transl Med* 2022;10(18):970. doi: 10.21037/atm-22-3901