Manuscript number HYPE201608906DR1

A proteomic-biostatistic integrated approach for finding the underlying molecular determinants of hypertension in human plasma

Prathibha R. Gajjala[1, 2)], Vera Jankowski[1)], Georg Heinze[3)], Grzegorz Bilo[4, 5)], Alberto Zanchetti[6, 7)], Heidi Noels[1)], Elisa Liehn[1)], Paul Perco[8)], Anna Schulz[9)], Christian Delles[10)], Felix Kork[1)], Erik Biessen[1,2)] , Krzysztof Narkiewicz[11)], Kalina Kawecka-Jaszcz[12)], Juergen Floege[13)], Davide Soranna[6,14)], Walter Zidek[9)], Joachim Jankowski[1, 2)]

1. Universitätsklinikum RWTH Aachen, Institute of Molecular Cardiovascular Research (IMCAR), Aachen, Germany
2. Experimental Vascular Pathology, Cardiovascular Research Institute Maastricht (CARIM), University of Maastricht, The Netherlands
3. Section for Clinical Biometrics, Center for Medical Statistics, Informatics and Intelligent Systems, Medical University of Vienna, Vienna, Austria
4. Department of Medicine and Surgery, University of Milano-Bicocca, Milan, Italy
5. Department of Cardiovascular, Neural and Metabolic Sciences, Istituto Auxologico Italiano, Milan, Italy
6. Istituto Auxologico Italiano, IRCCS, Milan, Italy
7. Università degli Studi di Milano, Milan, Italy
8. Department of Internal Medicine IV, Medical University Innsbruck, Innsbruck, Austria
9. Charité-Universitätsmedizin Berlin (CBF), Berlin, Germany
10. Institute of Cardiovascular and Medical Sciences, University of Glasgow, UK
11. Department of Hypertension and Diabetology, Medical University of Gdansk, Debinki 7c, 80-952 Gdansk, Poland
12. First Department of Cardiology, Interventional Electrocardiology and Hypertension, Jagiellonian University Medical College, Krakow, Poland
13. Internal Medicine II, Universitätsklinikum RWTH Aachen, Germany
14. Department of Statistics and Quantitative Methods, University of Milano-Bicocca, Milan, Italy

Short title: Plasma proteomics of hypertension

Address for correspondence
Univ.-Prof. Dr. Joachim Jankowski
Institute for Molecular Cardiovascular Research
RWTH Aachen University, University Hospital
Pauwelsstraße 30
D-52074 Aachen
Tel.: +49 (0) 241 80-80580
Fax: +49 (0) 241 80-82716
Email: jjankowski@ukaachen.de

Abstract

Despite advancements in lowering blood pressure, the best approach to lower it remains controversial due to the lack of information on the molecular basis of hypertension. We therefore performed plasma proteomics of plasma from patients with hypertension to identify molecular determinants detectable in these subjects but not in controls and vice versa. Plasma samples from hypertensive subjects (cases; n=118) and controls (n=85) from the “InGenious HyperCare” cohort were used for this study and performed mass spectrometric analysis. Using biostatistical methods, plasma peptides specific for hypertension were identified and a model was developed using least absolute shrinkage and selection operator logistic regression. The underlying peptides were identified and sequenced off-line using matrix assisted laser desorption ionization orbitrap mass spectrometry. By comparison of the molecular composition of the plasma samples, 27 molecular determinants were identified differently expressed in cases from controls. 70 % of the molecular determinants selected were found to occur less likely in hypertensive patients. In cross-validation, the overall R square was 0.434 and the area under the curve was 0.891 with 95% confidence interval 0.8482 to 0.9349, P<0.0001. The mean value of the cross-validated proteomic score of normotensive and hypertensive patients was found to be -2.007 ± 0.3568 and 3.383 ± 0.2643, P<0.0001 respectively. The molecular determinants were successfully identified and the proteomic model developed shows an excellent discriminatory ability between hypertensives and notmotensives. The identified molecular determinants may be the starting point for further studies to clarify the molecular causes of hypertension.

Key words: Hypertension, proteomics, blood pressure, anti-hypertensive drugs, modeling.

Introduction

A recent survey on Global Burden of Disease[1] has shown that, despite the availability of a number of effective blood pressure (BP) lowering drugs, the burden of disease caused by hypertension rather than decreasing, has continuously incremented worldwide. This indicates limitations of the diagnostic and therapeutic strategies so far implemented in managing hypertension and underscores the urgent need of new strategies for overcoming these limitations and reducing their consequences on public health.

Under the diagnostic aspect, hypertension is identified on the simple basis of a number of BP measurements taken under standard conditions, but it is well known that BP is extremely variable and its measurements are scarcely reproducible.[2] This has brought to a long lasting and continuing discussion on the preference to give to different BP assessment methods and settings: in the office by the doctor or the nurse, in the office with automatic equipment without the attendance of health professionals, at home or with 24-hour ambulatory monitoring. These different measurements also led to the definition of some categories or subtypes of hypertension such as “white coat hypertension” and “masked hypertension”.[2] Though the use of these subtypes of hypertension has become very popular in the management of hypertension, there are no agreement yet on strategies for their management.[3] Under the therapeutic aspect, a number of classes and compounds have been shown to effectively lower BP, thus reducing cardiovascular disease risk.[4, 5] However, each of these classes is known to be effective only in a proportion of hypertensive patients and in the absence of proven predictors of their effect they are commonly prescribed by a trial and error strategy, and often in association in order to more easily and promptly achieve therapeutic goal.[6]

A more precise diagnosis and a better-targeted treatment of hypertension may result from understanding of the genetic and molecular basis of hypertension. While a number of genome-wide association studies have been performed on large population samples or cohorts of hypertensives, the molecular definition of hypertension is largely unknown.[7] However, with advancements in the proteomics field, the molecular determinants at a low concentration i.e. femtomolar range in a complex biological specimen like plasma can be analyzed using mass-spectrometry. Several plasma peptides have been reported as relevant in hypertension.[8-11] Therefore, we investigated the differences in peptidic determinants profiles of plasma between normotensives and hypertensives by developing and applying a scoring system using a systems biology approach.

In addition, new systems-medicine based model for hypertension could be established and validated based on the identified molecular determinants of hypertension. The use of these models for prognosis of hypertension might be of relevance, because it is still unclear when to best initiate the treatment of hypertension.[6]

Methods

Study population

The “InGenious HyperCare” cohort (www.hypercare.eu) included individuals from families of probands with hypertension.[12] Among first-degree relatives of probands, at least one was hypertensive and one from a different generation. Hypertension was defined as office systolic blood pressure (SBP) ≥140 mm Hg or office diastolic blood pressure (DBP) ≥90 mm Hg or the presence of antihypertensive treatment and hypertension diagnosis before the age of 50. From the families enrolled in 4 centers participating in InGenious HyperCare gave a written informed consent and approved by local research ethics committees at each participating centre and selected 282

individuals for the current study. This number was subsequently reduced to 203 because of the lack of satisfactory blood samples in 79 cases. Phenotypic characterization included: basic anthropometric and clinical information, office BP and organ damage assessment. Office BP was measured with the auscultatory method (at least two seated measurements were obtained in standard conditions). Organ damage was defined as the presence of any of the following: 1) left ventricular hypertrophy on echocardiography (left ventricular mass index, LVMI, indexed to body surface area) ≥115 g/m^2 in males and ≥95 g/m^2 in females; echocardiographic images from all centers were analyzed centrally at Istituto Auxologico Italiano, Milan; 2) presence of microalbuminuria (albumin/creatinine ratio ≥ 3.4 mg/mmol; samples analyzed by local laboratories); 3) presence of chronic kidney disease (estimated glomerular filtration ratio (eGFR) by CKD-EPI formula <60 ml/min/1.73 m^2); 4) intima-media thickness of common carotid artery ≥0.9 mm on ultrasonography.

Proteomic analysis

Plasma proteomics was performed at the University Hospital RWTH Aachen, Institute for Molecular Cardiovascular Research (IMCAR) (Aachen, Germany) by using liquid chromatography-electrospray ionization-mass-spectrometry (LC-ESI-MS), matrix-assisted-laser-desorption ionization-time of flight-mass-spectrometry (MALDI-TOF-MS), MALDI LTQ Orbitrap XL and gas chromatography mass spectrometry (GC-MS). **Figure 1A** shows the workflow for proteomic analysis.

Statistical Methods

For statistical analysis we used Statistical Analysis System (SAS) software version 9.3 (SAS Institute, USA), Graph Pad Prism 6.0 software and R version 3.2.4.[13] Generally, continuous variables are reported as means and standard deviation (SDs) or median

and inter quartile range in case of non-normally distributed data and compared between groups using unpaired t-tests. Categorical variables are presented as percentages and frequencies. For plasma proteomic model development, we employed LASSO logistic and ridge regressions, flow chart of biostatistical analysis is shown in **Figure 2A**. More details about sample preparation and development of the model can be found in supplementary information of the manuscript.

In order to assess whether the cross-validated proteomic score was associated with clinical variables, we performed statistical tests (unpaired t-test or Mann-Whitney U test; Pearson's or Spearman's correlation) as well as multivariable logistic regression with hypertension status as binomial outcome and the cross-validated proteomic score, age, sex, BMI, diabetes, use of renin-angiotensin-aldosterone antagonists, SBP and DBP, heart rate and the presence of organ damage as explanatory variables. Additionally, multiple regression analysis with proteomic score as dependent variable and hypertension status, age, sex, BMI, diabetes, use of renin-angiotensin-aldosterone antagonists, SBP and DBP, heart rate and the presence of organ damage as predictor variables. We also compared the proteomic score according to hypertension control and the use of antihypertensive medication in the analysis of variance (ANOVA).

RESULTS

Characteristics of subjects

The study population was divided into 118 hypertensives (cases) and 85 controls according to the criteria described above. The characteristics of patients suffering from hypertension and controls are shown in **Table 1**. The percentage of males was 45.9% and 50.0% in controls and cases, respectively. Patients suffering from hypertension had

a SBP/DBP of 150.2±21.5/88.6±12.1 mm Hg compared to 131.0±14.9/79.7±9.9 mm Hg in the control group. The mean age, weight, body mass index and body surface area were found to be significantly higher in hypertensive patients as well (**Table 1**). Nine patients and three controls were diabetic. Coronary disease and myocardial infarction were diagnosed in 9 and 6 patients respectively, but not in controls. Furthermore, creatinine was higher and eGFR levels were significantly lower in hypertensive cases *vs.* normotensive subjects respectively. Significant differences were observed in glycemia and triglyceride levels and no differences were observed in lipid levels except for triglycerides. As expected cases had higher prevalence of organ damage overall as well as of each evaluated marker separately (**Table 1**).

Plasma proteomics

To identify the molecular determinants, we employed a plasma peptidomic approach by using mass-spectrometry. A characteristic total ion chromatogram (TIC) of a case and a control is shown in **Figure 1B** and **Figure 1C** respectively, showing the summation of the intensities of all mass signals detected as a function of LC retention time within a single run. The TICs of a case and a control were slightly different from each other. A corresponding representative average mass spectrum of a case and a control is depicted in **Figure 1D** and **Figure 1E**, respectively.

Biostatistical analysis

The raw data obtained from the mass-spectrometry were normalized using the internal standard. Then, an algorithm for peak picking was employed to combine all ions that derive from the same compound, thus considerably reducing the size and complexity of the dataset to be analyzed. Further, to simplify the data for statistical analysis, chromatograms were transferred into buckets with the information on intensity, m/z and

the retention time of the each molecular feature. Overall, 12,926 molecular features were detected, of which 403 features had 16.7% or more non-zero intensity values across all plasma samples. From these 403 features, 27 features were selected as predictive for the case-control status using LASSO logistic regression. Regression coefficients for the D (average non-zero vs. zero intensity) and X components (given non-zero intensity, log odds per tenfold increase of intensity) of selected peptides were ranked and depicted. The resulting proteomic model had a global p-value < 0.01 and an overall R^2 of 0.434. We found that down-regulation of 19 features and up-regulation of 8 features was associated with a higher probability of hypertension (**Table 2**). No additional molecular determinants were identified if model development was repeated after removing the features selected in the first run.

A proteomic score was calculated for each subject by cross-validation by leaving out a subject. **Figure 2B** depicts the proteomic scores as boxplots. The scores of hypertensive patients were higher when compared to the control subjects, with the mean values of the predictor score being 3.383 ± 0.2643 and -2.007 ± 0.3568, p<0.0001 respectively. Furthermore, the diagnostic power of these 27 molecular determinants was determined by receiver operating characteristic (ROC) analysis (sensitivity vs. 1-specificity, **Figure 2C**). The concordance index (area under the ROC) was 0.891 (95% confidence interval 0.8482 to 0.9349, P<0.0001).

Clinical correlates of the proteomic score

In univariable analyses the proteomic score correlated with age (r=0.43), BMI (r=0.22), SBP (r=0.29), DBP (r=0.16), creatinine (r=0.18), eGFR (r=-0.27), LVMI (corrected for body surface area – r=0.25 or for height$^{2.7}$ – r=0.28), IMT (r=0.24), urinary albumin/creatinine ratio (r=0.28) and the number of antihypertensive drugs (r=0.49). It

was higher in subjects with history of coronary disease (3.91±3.2 vs. 1.00±4.0, p=0.034) or myocardial infarction (5.08±2.7 vs. 1.01±4.0, p=0.015), in those with any organ damage (2.19±4.0 vs. 0.14±3.5, p=0.013), and a tendency toward higher values was also observed in males (1.62±4.0 vs. 0.67±4.1, p=0.095) and in diabetic subjects (3.23±4.34 vs. 0.96±4.0, p=0.058).

In multiple logistic regression analysis the association with the presence of hypertension was significant for the proteomic score, age and presence of organ damage (**Table 3**), with model R^2 = 0.65. A further multiple regression analysis was done with proteomic score as the dependent variable. In this model, the presence of hypertension was the only strong independent predictor, with BMI also showing borderline significant association; age was no longer significant in this model. The total predictive capacity of the model was R^2 =0.42; **Table 4**). Considering the subgroups of cases we found no significant difference in the proteomic score between treated and untreated (3.58±2.9 vs. 2.22±2.4, p=0.07), between uncontrolled and controlled individuals (3.2±3.0 vs. 3.80±2.3, p=0.34), between those with or without organ damage (2.96±3.4 vs. 3.51±2.7, p=0.38) and between those with or without any complication of hypertension (organ damage, diabetes, history of CVD) (3.30±3.2 vs. 3.38±2.7, p=0.90).

Identification of selected features and associated pathways

The 27 selected molecular determinants were identified using MALDI-TOF-TOF-MS and verified with LTQ-Orbitrap MS. **Figure 2D** represents a characteristic mass-spectrum of a plasma sample with indication of a molecular feature with m/z 736.9. Its respective fragmentation spectra is shown in **Figure 2E.** The signal at m/z 736.9 was found to be a fragment of phosphoinositide 3-kinase regulator (PI3KR1). The other features integrated in our final predictive hypertension model were identified as fragments of humanin (MT-

RNR2), ancotamin 10 (ANO10), NIK related protein kinase (NRK), mannose-6-phospho isomerase (MPI), tryptophan, erythrocyte membrane glycopeptide, transcription factor Dp-2 (TFDP2), pleckstrin homology domain-containing family (PLEKHO1), cardiac phospholamban (PLN), osteocalcin (BGLAP) or sarcolipin (SLN), ras-related protein Rab-13 (RAB13), protein prune homolog (PRUNE), nexilin (NEXN) and paladin (PALLD) proteins. One of the features was identified as tryptophan. The whole list of identified molecular determinants is presented in **Table S1** of supplementary data. To get an insight into pathophysiological role of these molecular determinants, we did literature mining. The functions and related pathways associated with the molecular determinants were summarized in **Table S1.** The pathway information was extracted from Gene Cards, and reported only the pathways, which had a score above 0.5. For PI3KR1, super pathways were reported, as this protein is associated with many pathways. In order to gain mechanistic insight into the pathophysiology of hypertension, KEGG and GO database searches were performed based on the 27 molecular determinants integrated in our proteomic model (supplementary data).

Discussion

In the present study, a comparative analysis of hypertensive patients *vs.* controls was performed on molecular level using systems biology approach. We developed a proteomic model based on the differences in peptidic profiles of plasma using 27 molecular determinants. The model shows an excellent discriminatory ability. Moreover, stability of the model is suggested since no additional molecular determinants could be detected if the originally selected determinants were omitted from model development. However, the low drop in R^2 values of the individual components of the combined proteomic score show that it is not a single molecular determinant that dominates the

model but rather the combination of many different molecular properties that is responsible for the superior performance of the model.

Using mass-spectrometry, we were able to identify 18 from the 27 molecular determinants selected, whereas the remaining 9 features remain currently unknown. Among the identified molecular determinants, 66.6% have negative β_{10fold} and $\beta_{nonzero}$ values, implying that they act as protective molecular determinants. Three proteins namely, humanin, osteocalcin and sarcolipin were included twice in the model with different signature sequence.

The peptidomic changes found in hypertensive patients may have several reasons: first these changes may be related to the pathogenesis of hypertension. We may assume such a relationship for those peptidic molecular determinants, which are known to mediate vasoconstriction or vascular smooth muscle cell growth.[14, 15] Second, the change in the peptidome of hypertensives may be a consequence of hypertension. This relationship may be assumed for peptides, which are secreted from myocardial cells. Increased amounts of these peptides may be secreted due to left ventricular hypertrophy.[16-18] Third, peptidomic changes may be the consequence of antihypertensive treatment. It is conceivable that especially substances blocking neurohumoral transmission, such as ACE inhibitors or beta-blockers, can induce counter regulatory processes.[19] These may also include the secretion of vasoregulatory peptides. It is beyond the scope of this study to review the functions of the peptides associated proteins extensively but these are briefly explained in **Table S1**. We identified peptide fragments of sarcolipin and phospholamban, which are proteins involved in cellular calcium metabolism and vessel contraction.[20, 21] Furthermore, we identified peptide fragments originating from humanin and osteocalcin proteins involved in

atherogenesis,[18, 22, 23] proteins involved in cytoskeletal organization and regulation (pleckstrin, palladin and nexilin), but also cellular proliferation is affected by protein prune homolog, NIK related protein kinase, transcription factor Dp-2.[24-26] Lastly, angiogenesis is a target of ras-related protein Rab-13.[27] Mannose phosphate isomerase is a ubiquitous enzyme, which, however, might be involved in the mannosylation of prorenin, thus facilitating its cellular uptake and hence angiotensinogen splitting.[28]

From the various physiological effects of these peptides it is difficult to decide whether changes in these peptides are related to the cause or the consequence of hypertension, or which are related to causes and which to consequences. On the one hand, altered angiogenesis, contraction or cytoskeleton may be pathogenetically important, but on the other hand hypertension may induce both left ventricular and vascular hypertrophy and, as a further consequence, also changes in peptidic messengers regulating cardiovascular structure and function. The proteomic approach focus on the identification of potential bioactive substances therefore, each of the substances identified and their combination in the current study should be characterized in in vivo models in future. Furthermore, future research has to be performed to determine if some of these molecules are present before the development of clinically detectable hypertension.

Multiple regression analyses revealed that hypertension was the strongest predictor of the proteomic score and this association was independent of age, sex, antihypertensive treatment, comorbidities or organ damage. This suggests that this score, when elevated, may indeed reflect the alterations of regulatory mechanisms leading to hypertension development or very early cardiac and vascular changes, rather than being the consequence of advanced hypertension. In this respect, the population of subjects

provided by the InGenious HyperCare cohort was particularly suitable for exploring mechanisms of hypertension, as hypertension was defined as high blood pressure diagnosed before age 50 years, thus excluding hypertension initiated at an older age when elevated blood pressure is largely dependent on large artery stiffening.

Several previous attempts have been made to elucidate hypertension-specific changes in the plasma peptidome. Araki et al.[29] studied the plasma peptidome from hypertensive pregnant women. In this study, 23 peptide peaks differed significantly between hypertensive pregnant women and healthy controls, with 11 peptides showing lower concentrations and 12 peptides showing higher concentrations in hypertension. Seven peptides were proteolytic fragments of higher molecular plasma proteins, suggesting an enhanced activity of proteolytic enzymes. Myers et al.[30] similarly performed mass spectrometric studies to identify proteins or peptides as markers of preeclampsia, which is a hypertensive pregnancy complication. They identified other markers than the aforementioned study, possibly due to methodological reasons. Gebhard et al.[31] showed that proteomic changes might also be the consequence of hypertension. They found that Ang II induced the biosynthesis of several cytoskeletal proteins in platelets. Other proteins were identified as pleckstrin and RAS-related protein Rab-11A. Similar peptides were also identified in the present study, suggesting that some of the changes described in our study may indeed be the consequence of high BP. Matafora et al.[32] followed a similar approach to determine changes in the urinary proteome specific for hypertension. They also identified proteomic changes, which were most likely markers of renal hypertensive damage.

Several limitations of the present study have to be mentioned: First, as detailed above, no conclusions as to the etiology of hypertension can be made from these data. A

causal relationship cannot be inferred from a statistical association, and influences from antihypertensive treatment may contribute to the peptidomic changes. Nevertheless, the present methodology may open up a new approach to define factors in the etiology of primary hypertension. Therefore, this study may stimulate new questions, which can be answered by studying selected groups of hypertensive patients and prehypertensive patients, e.g. those still untreated or those without vs. with left ventricular hypertrophy. Moreover, a study on secondary hypertensives may help understanding which peptidomic changes are rather a consequence than a cause of primary hypertension. Also, a study with longer follow-up may enable to identify molecular determinants and an associated proteomic-scoring model to predict the development of a comorbid disease status (e.g. CVD, diabetes) in hypertensive patients at later time points.

In conclusion, plasma analysis by mass spectrometry enabled us to short list a series of molecules linked to hypertensive patho-biology. Our findings demonstrated that with appropriate technologies plasma could be used as a source for the identification of molecular determinants of hypertension. Since hypertension is noted as one of the major risk factors of CVD, a similar approach could be used in subsequent studies to improve the screening and diagnosis of patients that are at a risk for CVD.

Acknowledgement

We thank Prof. Antonella Zambon for her support in the statistical analysis of the data.

Perspectives

In this case-control study, we investigated the differences in peptidic profiles of plasma from normotensives and hypertensives by developing and applying a scoring system using a systems biology approach. In addition, new systems-medicine based model for

hypertension is established and has an excellent discriminatory ability based on the identified molecular determinants of hypertension. In future, the substances identified in this study have to be validated as mediators of hypertension in animal studies and bioassays.

Sources of Funding

The work leading to this paper received funding from the European Community's Seventh Framework Programme under grant agreement no. HEALTH 2011.2.4.2-2" to "EU-MASCARA" (278249) and under grant agreement no.PEOPLE-ITN-GA-2013-608332 to "iMODE-CDK" from the European Union. We also received funding from "Corona foundation" under project number 5199/10062/2015.

Conflicts of Interest/Disclosures

No conflict of interest

References

1. Collaborators GBDRF, Forouzanfar MH, Alexander L et.al,: A systematic analysis for the global burden of disease study 2013. *Lancet*. 2015;386:2287-2323.
2. Stergiou GS, Parati G, Vlachopoulos C, et.al.. Methodology and technology for peripheral and central blood pressure and blood pressure variability measurement: Current status and future directions - position statement of the european society of hypertension working group on blood pressure monitoring and cardiovascular variability. *J Hypertens*. 2016;34:1665-1677.
3. Mancia G, Bombelli M, Seravalle G, Grassi G. Diagnosis and management of patients with white-coat and masked hypertension. *Nat Rev Cardiol*. 2011;8:686-693.

4. Thomopoulos C, Parati G, Zanchetti A. Effects of blood pressure lowering on outcome incidence in hypertension: 4. Effects of various classes of antihypertensive drugs--overview and meta-analyses. *J Hypertens*. 2015;33:195-211.
5. Thomopoulos C, Parati G, Zanchetti A. Effects of blood pressure-lowering on outcome incidence in hypertension: 5. Head-to-head comparisons of various classes of antihypertensive drugs - overview and meta-analyses. *J Hypertens*. 2015;33:1321-1341.
6. Mancia G, Fagard R, Narkiewicz K, et.al..2013 esh/esc guidelines for the management of arterial hypertension: The task force for the management of arterial hypertension of the european society of hypertension (esh) and of the european society of cardiology (esc). *J Hypertens*. 2013;31:1281-1357.
7. Ehret GB. Genome-wide association studies: Contribution of genomics to understanding blood pressure and essential hypertension. *Curr Hypertens Rep*. 2010;12:17-25.
8. Gamella-Pozuelo L, Fuentes-Calvo I, Gomez-Marcos MA, Recio-Rodriguez JI, Agudo-Conde C, Fernandez-Martin JL, Cannata-Andia JB, Lopez-Novoa JM, Garcia-Ortiz L, Martinez-Salgado C. Plasma cardiotrophin-1 as a marker of hypertension and diabetes-induced target organ damage and cardiovascular risk. *Medicine (Baltimore)*. 2015;94:e1218: 1-8.
9. Wang TJ, Gona P, Larson MG, Levy D, Benjamin EJ, Tofler GH, Jacques PF, Meigs JB, Rifai N, Selhub J, Robins SJ, Newton-Cheh C, Vasan RS. Multiple biomarkers and the risk of incident hypertension. *Hypertension*. 2007;49:432-438.

10. Kuklinska AM, Mroczko B, Musial WJ, Usowicz-Szarynska M, Sawicki R, Borowska H, Knapp M, Szmitkowski M. Diagnostic biomarkers of essential arterial hypertension: The value of prostacyclin, nitric oxide, oxidized-ldl, and peroxide measurements. *Int Heart J*. 2009;50:341-351.
11. Jan Danser AH. Renin and prorenin as biomarkers in hypertension. *Curr Opin Nephrol Hypertens*. 2012;21:508-514.
12. Neisius U, Bilo G, Taurino C, et.al.. Association of central and peripheral pulse pressure with intermediate cardiovascular phenotypes. *J Hypertens*. 2012;30:67-74.
13. R Development Core Team, R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, URL http://www.R-project.org, 2016.
14. Stewart DJ, Levy RD, Cernacek P, Langleben D. Increased plasma endothelin-1 in pulmonary hypertension: Marker or mediator of disease? *Ann Intern Med*. 1991;114:464-469.
15. Giaid A, Yanagisawa M, Langleben D, Michel RP, Levy R, Shennib H, Kimura S, Masaki T, Duguid WP, Stewart DJ. Expression of endothelin-1 in the lungs of patients with pulmonary hypertension. *N Engl J Med*. 1993;328:1732-1739.
16. Daniels LB, Maisel AS. Natriuretic peptides. *J Am Coll Cardiol*. 2007;50:2357-2368.
17. Horwich TB, Patel J, MacLellan WR, Fonarow GC. Cardiac troponin i is associated with impaired hemodynamics, progressive left ventricular dysfunction, and increased mortality rates in advanced heart failure. *Circulation*. 2003;108:833-838.

18. Guo B, Zhai D, Cabezas E, Welsh K, Nouraini S, Satterthwait AC, Reed JC. Humanin peptide suppresses apoptosis by interfering with bax activation. *Nature*. 2003;423:456-461.

19. Lazaro A, Gallego-Delgado J, Osende JI, Egido J, Vivanco F. Analysis of antihypertensive drugs in the heart of animal models: A proteomic approach. *Methods Mol Biol*. 2007;357:45-58.

20. MacLennan DH, Kranias EG. Phospholamban: A crucial regulator of cardiac contractility. *Nat Rev Mol Cell Biol*. 2003;4:566-577.

21. MacLennan DH, Asahi M, Tupling AR. The regulation of serca-type pumps by phospholamban and sarcolipin. *Ann N Y Acad Sci*. 2003;986:472-480.

22. Yen K, Lee C, Mehta H, Cohen P. The emerging role of the mitochondrial-derived peptide humanin in stress resistance. *J Mol Endocrinol*. 2013;50:R11-19.

23. Gossl M, Modder UI, Atkinson EJ, Lerman A, Khosla S. Osteocalcin expression by circulating endothelial progenitor cells in patients with coronary atherosclerosis. *J Am Coll Cardiol*. 2008;52:1314-1325.

24. Lemmon MA, Ferguson KM, Abrams CS. Pleckstrin homology domains and the cytoskeleton. *FEBS Lett*. 2002;513:71-76.

25. Dixon RD, Arneman DK, Rachlin AS, Sundaresan NR, Costello MJ, Campbell SL, Otey CA. Palladin is an actin cross-linking protein that uses immunoglobulin-like domains to bind filamentous actin. *J Biol Chem*. 2008;283:6222-6231.

26. Lee A, Hakuno F, Northcott P, Pessin JE, Rozakis Adcock M. Nexilin, a cardiomyopathy-associated f-actin binding protein, binds and regulates irs1 signaling in skeletal muscle cells. *PLoS One*. 2013;8:e55634: 1-10.

27. Gao J, Shen M, Guo X, Li X, Li J. Proteomic mechanism of myocardial angiogenesis augmented by remote ischemic training of skeletal muscle in rabbit. *Cardiovasc Ther*. 2011;29:199-210.

28. Re RN, Cook JL. Mechanisms of disease: Intracrine physiology in the cardiovascular system. *Nat Clin Pract Cardiovasc Med*. 2007;4:549-557.

29. Araki Y, Nonaka D, Tajima A, et.al.. Quantitative peptidomic analysis by a newly developed one-step direct transfer technology without depletion of major blood proteins: Its potential utility for monitoring of pathophysiological status in pregnancy-induced hypertension. *Proteomics*. 2011;11:2727-2737.

30. Myers JE, Tuytten R, Thomas G, Laroy W, Kas K, Vanpoucke G, Roberts CT, Kenny LC, Simpson NA, Baker PN, North RA. Integrated proteomics pipeline yields novel biomarkers for predicting preeclampsia. *Hypertension*. 2013;61:1281-1288.

31. Gebhard S, Steil L, Peters B, Gesell-Salazar M, Hammer E, Kuttler B, Clemetson KJ, Scharf C, Peters J, Volker U, Rettig R, Greinacher A. Angiotensin ii-dependent hypertension causes reversible changes in the platelet proteome. *J Hypertens*. 2011;29:2126-2137.

32. Matafora V, Zagato L, Ferrandi M, Molinari I, Zerbini G, Casamassima N, Lanzani C, Delli Carpini S, Trepiccione F, Manunta P, Bachi A, Capasso G. Quantitative proteomics reveals novel therapeutic and diagnostic markers in hypertension. *BBA Clin*. 2014;2:79-87.

Novelty and Significance

What Is New?

A more precise diagnosis and a better-targeted treatment of hypertension may result from understanding of the genetic and molecular basis of hypertension. Therefore, we analyzed plasma proteomic phenotypes of hypertension by robust proteo-biostatistic integrated approach with the “InGeniousHypercare” cohort, a cohort of hypertensive probands.

What Is Relevant?

By plasma peptidomic comparative analysis we identified the molecular features significantly different in hypertensive and control subjects. These features provide new insights in the genesis and progression of hypertension and may provide new targets for the treatment of hypertension.

Summary

In the current systems medicine based study we were able to identify significant differently expressed peptides and proteins in subjects with hypertension and control subjects. These molecular features will be useful to clarify the molecular causes of hypertension and to predict the development of hypertension and of associated cardiovascular events in the future.

Figure legends

Figure 1: Sample preprocessing and mass spectrometric analysis

(A) Outline of the steps followed in sample preparation for mass spectrometer and processing the data for statistical analysis

(B) Characteristic total ion chromatogram of a sample from the case subgroup

(C) Characteristic total ion chromatogram of a sample from the control subgroup

(D) Corresponding average mass spectrum of a sample from the case subgroup

(E) Corresponding average mass spectrum of a sample from the control subgroup

Figure 2: Development of a predictor model and identification of peptides that distinguish the hypertensive and normotensive subjects

(A) Schematic representation of steps involved in the development of a predictor model for hypertension

(B) Box-plot of a cross-validated plasma proteomics model for hypertension, $p<0.0001$

(C) ROC curve of the cross-validated plasma proteomics model with an area under curve of 0.891 (95% confidence interval 0.8482 to 0.9349. $P<0.0001$)

(D) Representative mass spectrum of the selected molecular feature with m/z of 736.9

(E) Representative fragmentation spectra of the selected molecular feature with m/z of 736.9

Table 1. Baseline characteristics of InGenious HyperCare cohort

Variable	Controls (n=85)	Cases (n=118)	P-value
Demographics			
Male, n (%)	39 (45.9)	59 (50.0)	n.s
Age (years)*	34 (27-45)	53 (48-60)	< 0.0001
Physical examination			
Height (cm)	170.7 ± 9.8	169.0 ± 9.1	n.s
Weight (kg)	77.3 ± 13.7	86.9 ± 16.3	< 0.0004
Body mass index (kg/m2)	26.6 ± 4.4	30.4 ± 4.9	< 0.0001
Body surface area (m2)	1.9 ± 0.2	2.0 ± 0.2	0.0222
SBP (mmHg)*	133.0 (123.0-138.0)	148.0 (135.8-162.0)	< 0.0001
DBP (mmHg)*	79.0 (73.0-86.0)	88.0 (81.0, 95.3)	< 0.0001
Heart rate (bpm)	70.1 ± 9.9	68.9 ± 11.0	n.s
Clinical history			
Diabetes, n (%)	3 (3.5)	9 (7.6)	n.s
Coronary disease, n (%)	0	9 (7.6)	0.0066
Myocardial infarction, n (%)	0	6 (5.0)	0.0365
Medications			
Antihypertensive treatment, n (%)	8 (9.4)	111 (94.1)	<0.0001
Use of RAAS blocker, n (%)	5 (5.6)	82 (69.5)	<0.0001
Biochemical data			
Creatinine (µmol/L)	72.9 ± 10.9	79.3 ± 26.7	0.0392
eGFR (ml/min)	93.5 ± 19.7	80.3 ± 16.7	< 0.0001
Glycemia (mmol/L)*	4.8 (4.5-5.3)	5.3 (4.8-5.9)	0.0002
Total cholesterol (mmol/L)	5.2 ± 1.1	5.2 ± 1.2	n.s
HDL (mmol/L)	1.4 ± 0.3	1.4 ± 0.7	n.s
LDL (mmol/L)	3.2 ± 0.9	3.0 ± 1.1	n.s
Triglyceride (mmol/L)	1.5 ± 1.4	1.8 ± 0.9	0.0006
Organ damage			

Left ventricular mass index (g/m2)	80.9 ± 1.6	92.3 ± 1.8	< 0.0001
Left ventricular mass index (g/m2.7)	36.3 ± 0.9	44.2 ± 1.0	< 0.0001
Left ventricular hypertrophy, N(%)	8 (9.5)	26 (22.6)	0.0114
Albumin/creatinine ratio (mg/mmol) *	0.48 (0.00-0.89)	1.15 (0.56-2.50)	0.0001
Microalbuminuria, N(%)	1 (2.3)	11 (20.0)	0.0062
Chronic kidney disease, N(%)	3 (3.6)	10 (8.9)	n.s.
Common carotid IMT (mm)	0.58 ± 0.12	0.68 ± 0.14	< 0.0001
IMT ≥0.9 mm, N(%)	3 (3.5)	11 (9.5)	0.0845
Any organ damage, N(%)	4 (4.7)	27 (22.9)	0.0002

* - median and interquartile range DBP=diastolic blood pressure, eGFR=estimated glomerular filtration rate, HDL=high density lipoprotein, IMT=intima-media thickness, LDL=low density lipoprotein, RAAS=renin angiotensin aldosterone system, SBP=systolic blood pressure, n.s=not significant**.**

Table 2. Identified features and their coefficients

m/z	β_{10fold}	$\beta_{nonzero}$	$Pr_{nonzero}$	Dynamic range	Drop in R^2
Down-regulated in hypertensive patients					
344.21	-0.9862	-1.2175	35	4.7940	-0.0006
811.37	-0.7529	-0.9328	35	1.6952	0.0023
358.23	-0.6049	-0.7483	41	2.7697	0.0032
1043.44	-0.4497	-0.7334	42	3.6804	0.0143
549.78	-0.6083	-0.7227	45	4.8103	0.0027
148.42	-0.499	-0.7150	54	6.0856	0.0354
317.24	-0.4611	-0.6449	54	3.0719	0.0002
173.30	-0.4507	-0.5295	58	4.9854	0.0003
346.02	-0.3984	-0.5177	58	3.9033	0.0029
167.27	-0.4534	-0.4719	59	3.3400	0.0008

803.25	-0.3249	-0.4650	61	10.1504	0.0344
1003.47	-0.3027	-0.4608	63	3.4437	0.0085
327.22	-0.4421	-0.4482	63	3.6884	0.0011
149.59	-0.4645	-0.4399	64	5.2870	0.0128
187.58	-0.4246	-0.4189	65	6.1546	0.0065
736.91	-0.3727	-0.3334	68	4.0306	0.0020
359.21	-0.1891	-0.2553	68	2.3112	0.0088
647.21	-0.1281	-0.2045	77	3.5131	0.0041
266.09	-0.1045	-0.0993	90	3.9237	-0.0003
Upregulated in hypertensive patients					
266.16	0.1246	0.1655	93	2.3868	0.0003
302.22	0.1856	0.2232	114	4.9577	0.0016
355.02	0.3620	0.3860	126	3.8948	0.0054
274.15	0.4822	0.5021	162	5.3861	0.0007
909.49	0.3429	0.5549	168	7.2189	0.0073
169.59	0.5351	0.9185	177	7.1732	0.0118
852.52	0.8124	1.3102	182	7.4410	0.0095
268.23	0.6820	1.3813	194	5.6759	0.0025

β_{10fold} and $\beta_{nonzero}$ describe the logistic regression coefficients (log odds increase for hypertension) per 10fold increase of an intensity value and for comparing average non-zero intensity to a zero intensity value, respectively. $Pr_{nonzero}$ denotes the proportion of nonzero values across the 203 samples. The dynamic range of intensity values refers to range on the log2 scale. i.e. how often the lowest nonzero intensity value has to be 'doubled' in order to be equal to the highest nonzero intensity value. Drop in R^2 describes the importance of each predictor in terms of drop in R^2 if the variable is remove

Table 3. Logistic regression model linking the presence of hypertension with the proteomic score and major clinical variables

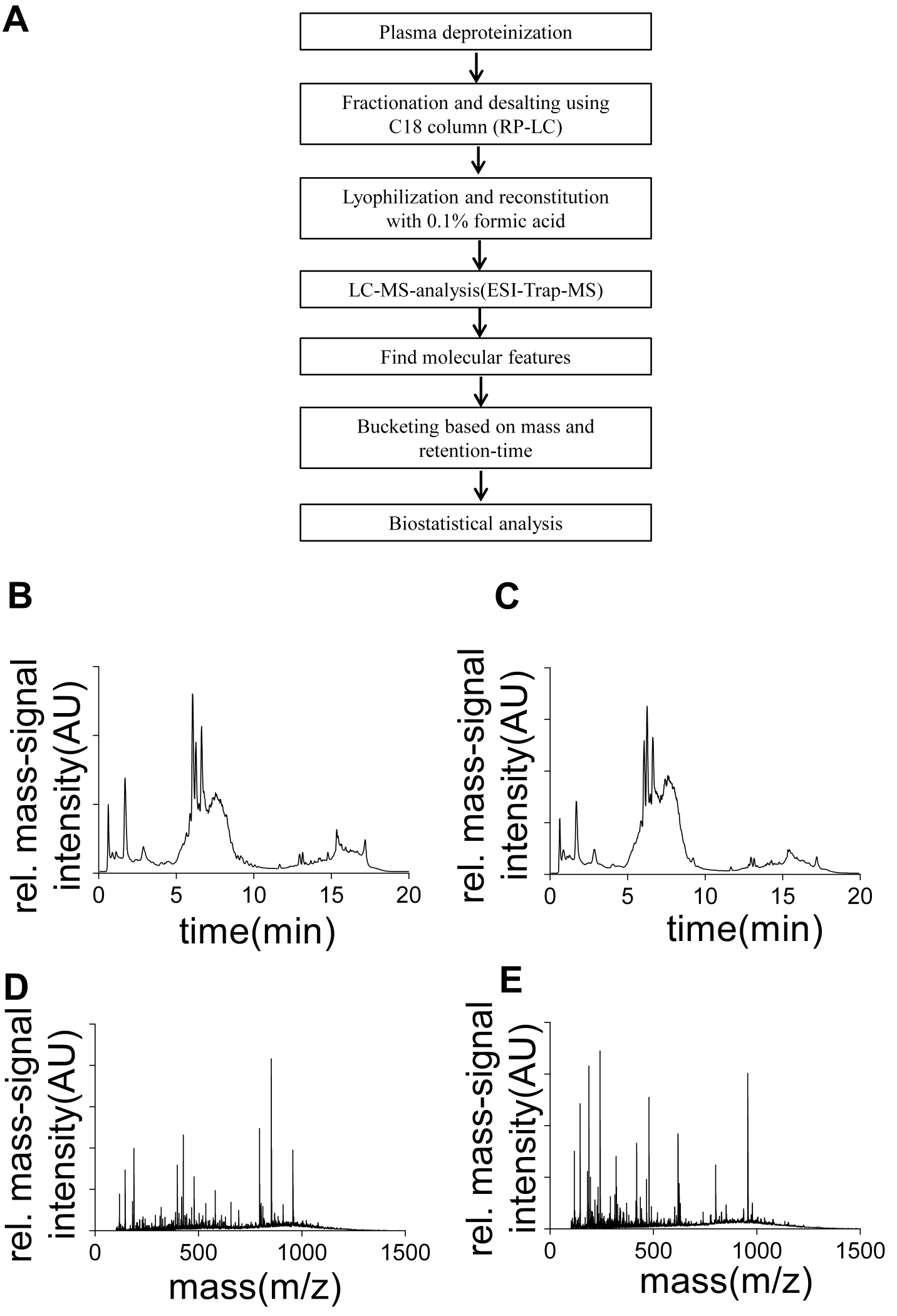
Variable	OR (95% CI)	P-value
Proteomic score	2.02 (1.58-2.57)	<0.0001
Age (per 10 years)	3.99 (2.25-7.07)	<0.0001
BMI	1.10 (0.98-1.23)	0.0915
Sex (Male vs Female)	1.66 (0.55-5.00)	0.3697
Organ Damage	13.68 (1.83-102.09)	0.0107
HR (per 10 bpm)	1.29 (0.70-2.38)	0.4148

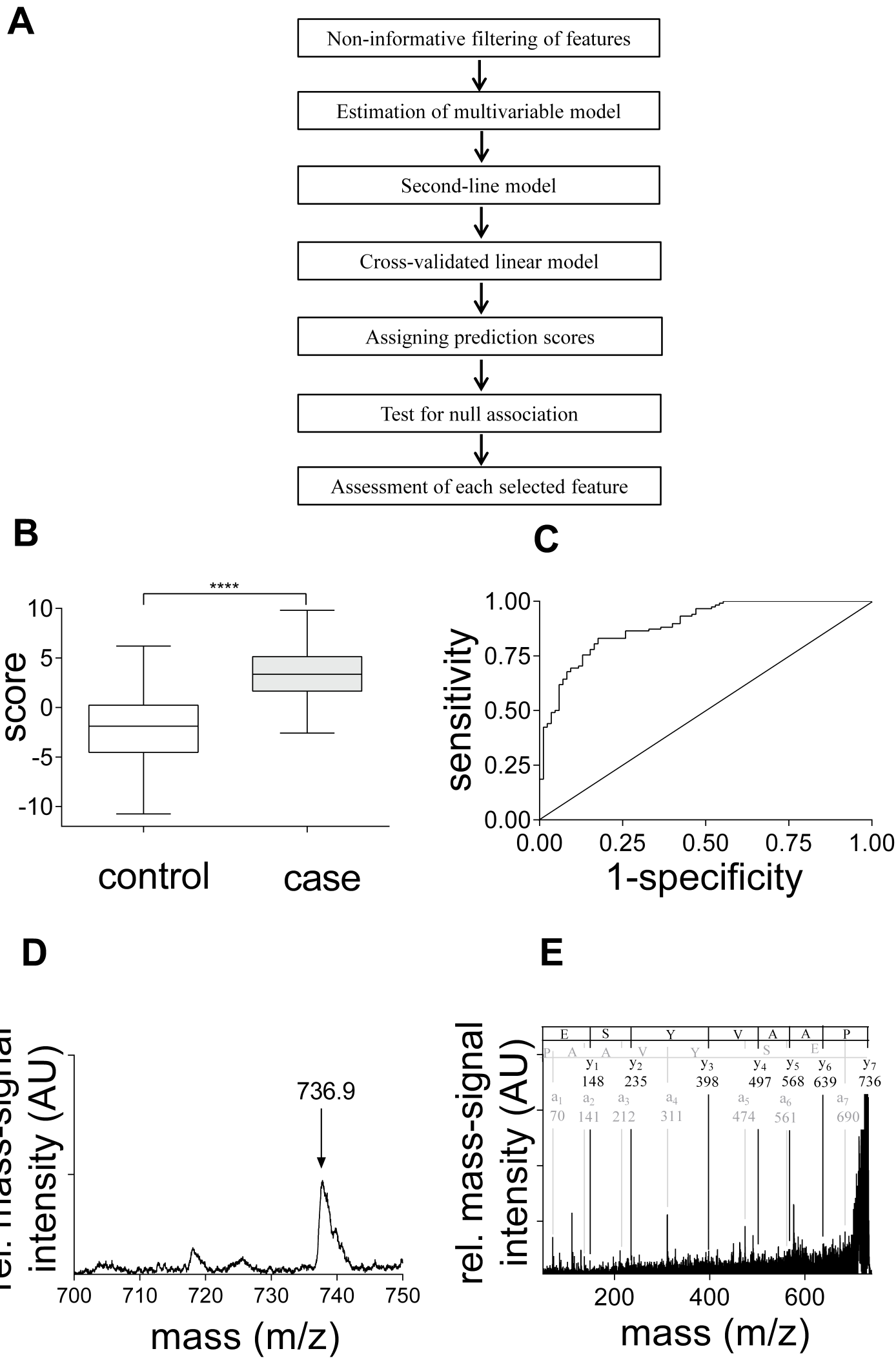
BMI=body mass index, HR=heart rate.

Table 4. Logistic regression model linking the proteomic score with hypertension and major clinical variables

Variables	Beta	SE of Beta	P-value
Sex	-0.181085	0.091495	0.051280
Age	-0.036913	0.134954	0.785167
hypertension	-0.624723	0.133414	0.000012
BMI	-0.221700	0.106824	0.041206
Organ Damage	0.039269	0.134910	0.771753
eGFR	-0.086320	0.101515	0.397721
UACR	0.163188	0.097648	0.098642
LVMI (Height)	0.075009	0.135947	0.582677
Glycemia	-0.105117	0.088147	0.236624

BMI=body mass index, eGFR=estimated glomerular filtration rate, LVMI=left ventricular mass index, SE=standard error, UACR=urinary albumin-creatinine ratio.


A
Plasma deproteinization
Fractionation and desalting using C18 column (RP-LC)
Lyophilization and reconstitution with 0.1% formic acid
LC-MS-analysis(ESI-Trap-MS)
Find molecular features
Bucketing based on mass and retention-time
Biostatistical analysis
B
rel. mass-signal intensity(AU)
0
5
10
15
20
time(min)
C
rel. mass-signal intensity(AU)
0
5
10
15
20
time(min)
D
rel. mass-signal intensity(AU)
0
500
1000
1500
mass(m/z)
E
rel. mass-signal intensity(AU)
0
500
1000
1500
mass(m/z)


A
Non-informative filtering of features
Estimation of multivariable model
Second-line model
Cross-validated linear model
Assigning prediction scores
Test for null association
Assessment of each selected feature
B
****
score
10
5
0
-5
-10
control
case
C
sensitivity
1.00
0.75
0.50
0.25
0.00
0.00
0.25
0.50
0.75
1.00
1-specificity
D
rel. mass-signal intensity (AU)
736.9
700
710
720
730
740
750
mass (m/z)
E
rel. mass-signal intensity (AU)
E S Y V A A P
y1 148
y2 235
y3 398
y4 497
y5 568
y6 639
y7 736
a1 70
a2 141
a3 212
a4 311
a5 474
a6 561
a7 690
200
400
600
mass (m/z)