Received: 11 June 2021 | Accepted: 14 December 2021

DOI: 10.1111/jth.15623

ORIGINAL ARTICLE

# Targeted proteomics for evaluating risk of venous thrombosis following traumatic lower-leg injury or knee arthroscopy

Yassene Mohammed[1,2] | Carolina E. Touw[3,4] | Banne Nemeth[3,4] | Raymond A. van Adrichem[3,4] | Christoph H. Borchers[5,6,7] | Frits R. Rosendaal[3] | Bart J. van Vlijmen[8] | Suzanne C. Cannegieter[3,8]

[1]Center for Proteomics and Metabolomics, Leiden University Medical Center, Leiden, The Netherlands

[2]University of Victoria - Genome British Columbia Proteomics Centre, Victoria, British Columbia, Canada

[3]Department of Clinical Epidemiology, Leiden University Medical Center, Leiden, The Netherlands

[4]Department of Orthopaedic Surgery, Leiden University Medical Center, Leiden, The Netherlands

[5]Segal Cancer Proteomics Centre, Segal Cancer Centre, Lady Davis Institute, Jewish General Hospital, McGill University, Montreal, Quebec, Canada

[6]Gerald Bronfman Department of Oncology, Jewish General Hospital, McGill University, Montreal, Quebec, Canada

[7]Department of Data Intensive Science and Engineering, Skolkovo Institute of Science and Technology, Skolkovo Innovation Center, Moscow, Russia

[8]Einthoven Laboratory for Experimental Vascular Medicine, Department of Internal Medicine, Division of Thrombosis & Hemostasis, Leiden University Medical Center, Leiden, The Netherlands

**Correspondence**
Yassene Mohammed, Center for Proteomics and Metabolomics, Leiden University Medical Center, P.O. Box 9600, 2300 RC Leiden, The Netherlands.
Email: y.mohammed@lumc.nl

Suzanne C. Cannegieter, Department of Clinical Epidemiology and Einthoven Laboratory for Vascular and Regenerative Medicine, Leiden University Medical Center, P.O. Box 9600, 2300 RC Leiden, The Netherlands.
Email: s.c.cannegieter@lumc.nl

**Funding information**
Genome British Columbia, Grant/Award Number: 204PRO, 214PRO, 264PRO and 282PQP; Genome Canada, Grant/Award Number: 204PRO, 214PRO, 264PRO and 282PQP

## Abstract

**Introduction:** Patients with lower-leg cast immobilization and patients undergoing knee arthroscopy have an increased risk of venous thrombosis (VT). Guidelines are ambiguous about thromboprophylaxis use, and individual risk factors for developing VT are often ignored. To assist in VT risk stratification and guide thromboprophylaxis use, various prediction models have been developed. These models depend largely on clinical factors and provide reasonably good C-statistics of around 70%. We explored using protein levels in blood plasma measured by multiplexed quantitative targeted proteomics to predict VT. Our aim was to assess whether a VT risk prediction model based on absolute plasma protein quantification is possible.

**Methods:** We used internal standards to quantify proteins in less than 10 µl plasma. We measured 270 proteins in samples from patients scheduled for knee arthroscopy or with lower-leg cast immobilization. The two prospective POT-(K)CAST trails allow complementary views of VT signature in blood, namely pre and post trauma, respectively. From approximately 3000 patients, 31 patients developed VT who were included and matched with double the number of controls.

Manuscript Handled by: Patricia Liaw

Final decision: Patricia Liaw, 14 December 2021

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.
© 2021 The Authors. *Journal of Thrombosis and Haemostasis* published by Wiley Periodicals LLC on behalf of International Society on Thrombosis and Haemostasis.

**Results:** Top discriminating proteins between cases and controls included APOC3, APOC4, APOC2, ATRN, F13B, and F2 in knee arthroscopy patients and APOE, SERPINF2, B2M, F13B, AFM, and C1QC in patients with lower-leg cast. A logistic regression model with cross-validation resulted in C-statistics of 88.1% (95% CI: 85.7–90.6%) and 79.6% (95% CI: 77.2–82.0%) for knee arthroscopy and cast immobilization groups respectively.

**Conclusions:** Promising C-statistics merit further exploration of the value of proteomic tests for predicting VT risk upon additional validation.

**KEYWORDS**

blood proteins, prediction, proteomics, quantitative evaluation, reference standards, venous thrombosis

## 1 | INTRODUCTION

Deep vein thrombosis (DVT) and pulmonary embolism (PE) are two manifestations of venous thrombosis (VT), which is associated with considerable mortality and morbidity.[1-3] Several genetic and environmental factors play a role in the development of VT.[4,5] VT is in general characterized with an incidence rate of 0.1–0.2% per year in the general population.[1] Patients with lower-leg cast immobilization and patients undergoing knee arthroscopy have a 56-fold and 16-fold increased risk of developing VT respectively, in the first three months compared with the general population.[6,7]

Recent studies have shown that thromboprophylactic treatment with low-molecular-weight heparin (LMWH) does not prevent VT following lower-leg casting or during the first 8 days after knee arthroscopy.[8] Despite LMWH treatment, 1.4% and 0.7% still develops VT, respectively, which was not lower than the incidence in the control groups.[8] These patients must have a strong propensity to develop VT to which thromboprophylactic treatment may need to be adjusted to either longer duration or higher dosage. For this, estimating an individual's VT risk is required, for which several prediction models have proven valuable (Table S1).[9-14] These models mostly depend on clinical information and patient history.[15-17] We assessed whether the prediction can be improved using protein absolute concentrations measured in blood.

We used targeted mass spectrometry-based proteomics to quantify blood plasma proteins and studied the association with subsequent VT in lower-leg casting and knee arthroscopy patients.[18-21] Our approach is multiplexed and requires relatively low amount/volume of sample, typically less than 10 μL plasma and over 100 target proteins.[22,23] The purpose of this study was to assess the possibility of using precise plasma protein abundances in predicting venous thrombosis after knee arthroscopy or after lower leg injury.

### ESSENTIALS

- Venous thrombosis (VT) risk prediction depends largely on clinical factors and allows reasonably good C-statistics of around 70%.
- Quantitative mass spectrometry-based proteomics with internal standard allows measuring absolute protein concentration in multiplex using low sample volume.
- 167 proteins were quantified in 10 μl plasma from two prospective cohorts with complementary views of VT signature in blood, namely pre and post trauma.
- Using cross validation and bootstrapping, discriminating proteins predicted VT risk with C-statistics >78% demonstrating a potential for proteomics VT prediction test.

## 2 | METHODS

### 2.1 | Patients and sample collection

We used samples acquired from two large multicenter randomized clinical trials, designed to study the effectiveness of thromboprophylaxis for prevention of symptomatic VT, i.e., the Prevention of Thrombosis after Lower Leg Plaster Cast (POT-CAST) trial and the Prevention of Thrombosis after Knee Arthroscopy (POT-KAST) trial. Study details have been described elsewhere[8] and the general characteristics of participants are summarized in Table 1. In short, between March 2012 and January 2016, patients aged 18 years or older, admitted to the emergency department who were treated with lower-leg cast immobilization (POT-CAST trial), and patients aged 18 years or older who had an indication for knee arthroscopy – such as meniscectomy, removal of loose bodies and diagnostic arthroscopy – were included (POT-KAST trial).

**TABLE 1** General characteristics of patients included in the current study to evaluate the protein signature of VTE

| | Knee arthroscopy | | Lower-leg injury | |
|---|---|---|---|---|
| | Cases (*N* = 8) | Controls (*N* = 16) | Cases (*N* = 23) | Controls (*N* = 46) |
| Sex | | | | |
| Male, *n* (%) | 4 (50.0) | 8 (50.0) | 13 (56.5) | 26 (56.5) |
| Female, *n* (%) | 4 (50.0) | 8 (50.0) | 10 (43.5) | 20 (43.5) |
| Age | | | | |
| Median in years (25th–75th percentile) | 50.5 (50.0–58.8) | 51.0 (49.0–60.3) | 55.1 (46.3–59.1) | 53.4 (45.1–60.7) |
| BMI | | | | |
| Median in years (25th–75th percentile) | 28.7 (25.7–31.2) | 27.0 (24.6–30.2) | 27.9 (24.5–30.3) | 26.4 (24.9–30.1) |
| Comorbidity | | | | |
| Yes, *n* (%) | 1 (12.5) | 2 (12.5) | 6 (72.7) | 5 (11.4) |
| No, *n* (%) | 7 (87.5) | 14 (87.5) | 16 (27.3) | 39 (88.6) |
| Smoking | | | | |
| Current smoker, *n* (%) | 0 (0.0) | 3 (18.8) | 9 (42.9) | 12 (27.3) |
| Non-smoker, *n* (%) | 2 (25.0) | 11 (68.8) | 5 (23.8) | 12 (27.3) |
| Former smoker, *n* (%) | 6 (75.0) | 2 (12.5) | 7 (33.3) | 20 (45.5) |
| Current use of oral contraceptives | | | | |
| Yes, *n* (%) | 1 (12.5) | 1 (6.3) | 2 (9.1) | 1 (2.3) |
| No, *n* (%) | 7 (87.5) | 15 (93.8) | 20 (90.9) | 43 (97.7) |
| Malignancy in last 5 years | | | | |
| Yes, *n* (%) | 1 (12.5) | 0 (0.0) | 0 (0.0) | 0 (0.0) |
| No, *n* (%) | 7 (87.5) | 16 (100.0) | 22 (100.0) | 44 (100.0) |
| Surgery in past 2 months | | | | |
| Yes, *n* (%) | 0 (0.0) | 1 (6.3) | 6 (28.6) | 4 (9.1) |
| No, *n* (%) | 8 (100.0) | 15 (93.8) | 15 (71.4) | 40 (90.9) |
| Immobility in last 2 months (for at least 4 days) | | | | |
| Yes, *n* (%) | 1 (12.5) | 0 (0.0) | 3 (14.3) | 1 (2.3) |
| No, *n* (%) | 7 (87.5) | 16 (100.0) | 18 (85.7) | 43 (97.7) |
| VT in first degree relatives | | | | |
| Yes, *n* (%) | 2 (25.0) | 0 (0.0) | 4 (21.1) | 5 (12.5) |
| No, *n* (%) | 6 (75.0) | 16 (100.0) | 15 (78.9) | 35 (87.5) |
| Varicose veins | | | | |
| Yes, *n* (%) | 2 (25.0) | 5 (31.3) | 4 (19.0) | 4 (9.1) |
| No, *n* (%) | 6 (75.0) | 11 (68.8) | 17 (81.0) | 40 (90.9) |
| ABO-blood type[a] | | | | |
| Homozygote O, *n* (%) | 3 (37.5) | 8 (50.0) | 7 (30.4) | 19 (45.2) |
| Heterozygote O, *n* (%) | 4 (50.0) | 7 (43.8) | 13 (56.5) | 21 (50.0) |
| Homozygote non-O, *n* (%) | 1 (12.5) | 1 (6.3) | 3 (13.0) | 2 (4.8) |
| FII20210 prothrombin mutation | | | | |
| No polymorphism, *n* (%) | 8 (100.0) | 16 (100.0) | 22 (95.7) | 42 (100.0) |
| Heterozygote, *n* (%) | 0 (0.0) | 0 (0.0) | 1 (4.3) | 0 (0.0) |
| Factor V Leiden | | | | |
| No polymorphism, *n* (%) | 7 (87.5) | 16 (100.0) | 19 (82.6) | 41 (97.6) |
| Heterozygote, *n* (%) | 1 (12.5) | 0 (0.0) | 4 (17.4) | 1 (2.4) |

(Continues)

TABLE 1 (Continued)

| | Knee arthroscopy | | Lower-leg injury | |
|---|---|---|---|---|
| | Cases (*N* = 8) | Controls (*N* = 16) | Cases (*N* = 23) | Controls (*N* = 46) |
| rs2066865 | | | | |
| No polymorphism, *n* (%) | 4 (50.0) | 7 (43.8) | 11 (47.8) | 22 (52.4) |
| Heterozygote, *n* (%) | 3 (37.5) | 6 (37.5) | 10 (43.5) | 18 (42.9) |
| Homozygote, *n* (%) | 1 (12.5) | 3 (18.8) | 2 (8.7) | 2 (4.8) |
| rs2289252 | | | | |
| No polymorphism, *n* (%) | 3 (37.5) | 3 (18.8) | 5 (21.7) | 20 (46.5) |
| Heterozygote, *n* (%) | 4 (50.0) | 10 (62.5) | 14 (60.9) | 17 (39.5) |
| Homozygote, *n* (%) | 1 (12.5) | 3 (18.8) | 4 (17.4) | 6 (14.0) |
| LMWH assignment | | | | |
| Yes, *n* (%) | 5 (62.5) | 7 (43.8) | 9 (42.9) | 20 (45.5) |
| No, *n* (%) | 3 (37.5) | 9 (56.3) | 12 (57.1) | 24 (54.5) |

[a]ABO-blood type is stratified based on number of alleles (genes): i.e. in case of 'homozygote O' there are two alleles of blood type O, in case of 'heterozygote O' there is one allele of blood type O and one allele of a blood type other than O (A or B), in case of 'homozygote non-O' there are two alleles of blood type other than O.

Each trial included approximately 1500 patients, randomized to either low molecular weight heparin or no treatment (ratio 1:1). All patients completed a questionnaire on risk factors for VT at time of inclusion. In patients with lower leg injury, blood was drawn upon presentation at the emergency department, and in knee arthroscopy patients, blood was drawn within 4 hours before they underwent surgery. Citrate plasma and EDTA plasma was collected, centrifuged at 2500 *g* for 10 min, and then transferred to micro centrifuge tubes. All samples were stored at −80°C within 4 hours after blood collection. The medical ethics committee at Leiden University Medical Center approved both trials (Table 2).

## 2.2 | Study design

In a case-control design, nested in these trials, cases were patients who developed VT within 3-months following inclusion (31 patients in total, i.e., 23 from the POT-CAST and 8 from the POT-KAST trial). Controls (ratio 1:2), were patients who did not develop VT, i.e., 46 from the POT-CAST and 16 patients from the POT-KAST trial (Figure 1). Controls were matched on age and sex, and to account for sampling storage variation, on the hospital of inclusion and the time and date of sample storage (in this order). For all analyses, EDTA samples were used. For a sensitivity analysis, for validation and comparison with EDTA samples, we also included 10 citrate plasma samples (from POT-CAST and POT-KAST patients, with and without venous thrombosis).

## 2.3 | Sample preparation

Targeted mass spectrometry-based proteomics assays were developed previously for 270 proteins covered by 274 surrogate peptides (Table S2). The assays development is part of our efforts to build a documented library of Multiple Reaction Monitoring (MRM) assays for blood plasma proteins.[22,24,25] The peptides for the MRM assays were selected by PeptidePicker[25] and synthesized and validated as previously described.[26,27] The sample preparation was performed in an automated manner using a Tecan Freedom Evo 150 robot. The Urea-based preparation protocol used was developed and applied previously.[28] For the current study, the digests were prepared through simultaneous denaturation and reduction of the homogenate with 9 M urea/20 mM dithiothreitol for 30 min at 37°C. Denatured proteins were alkylated with iodoacetamide (40 mM final concentration) for 30 min at room temperature, and then samples were diluted to reach a final urea concentration of 0.55 mM prior to tryptic digestion. Digestion was carried out at a 10:1 substrate:enzyme ratio using Tosyl phenylalanyl chloromethyl ketone –TPCK-treated trypsin (Worthington) for 18 h at 37°C. After digestion, samples were acidified with aqueous 1% formic acid (FA), and a chilled heavy labeled synthesized internal standard (SIS) peptide mixture was added. Samples were concentrated *via* solid phase extraction (10 mg Oasis HLB cartridges; Waters), using the manufacturer's recommended protocol. Finally, the samples were eluted with 55% acetonitrile (ACN)/0.1% FA (300 µl) and lyophilized to dryness. The dried samples were rehydrated in 0.1% formic acid to a final concentration of 1 µg/µl for LC/MRM-MS analysis.

## 2.4 | Targeted proteomics

The samples were separated on-line with a RP-UHPLC column (EclipsePlusC18 RRHD 150 x 2.1 mm i.d., 1.8 µm particle diameter; Agilent) maintained at 50°C. Peptide separations were performed at 0.4 ml/min over a 56 min run, *via* a multi-step LC gradient (1.5–81% 2–80%mobile phase B; mobile phase B: 0.1% FA in ACN). The

TABLE 2 Top discriminating proteins between VTE cases and controls following traumatic lower-leg injury or knee arthroscopy

| Protein names | UniprotKB accession | Gene names | Increase or decrease in cases relatively to controls |
|---|---|---|---|
| Knee arthroscopy | | | |
| Apolipoprotein C-III (Apo-CIII) (ApoC-III) (Apolipoprotein C3) | P02656 | APOC3 | Decrease |
| Attractin (DPPT-L) (Mahogany homolog) | O75882 | ATRN | Decrease |
| Apolipoprotein C-IV (Apo-CIV) (ApoC-IV) (Apolipoprotein C4) | P55056 | APOC4 | Decrease |
| Apolipoprotein C-II (Apo-CII) (ApoC-II) (Apolipoprotein C2) [Cleaved into: Proapolipoprotein C-II (ProapoC-II)] | P02655 | APOC2 | Decrease |
| Coagulation factor XIII B chain (Fibrin-stabilizing factor B subunit) (Protein-glutamine gamma-glutamyltransferase B chain) (Transglutaminase B chain) | P05160 | F13B | Decrease |
| Prothrombin (EC 3.4.21.5) (Coagulation factor II) [Cleaved into: Activation peptide fragment 1; Activation peptide fragment 2; Thrombin light chain; Thrombin heavy chain] | P00734 | F2 | Decrease |
| Plaster cast | | | |
| Apolipoprotein B–100 | P04114 | APOB | Increase |
| Apolipoprotein E (Apo-E) | P02649 | APOE | Increase |
| Alpha–2-antiplasmin (Alpha–2-AP) (Alpha–2-plasmin inhibitor) (Alpha–2-PI) (Serpin F2) | P08697 | SERPINF2 | Increase |
| Beta–2-microglobulin [Cleaved into: Beta–2-microglobulin form pI 5.3] | P61769 | B2 M | Increase in female, steady in male |
| Afamin (Alpha-albumin) (Alpha-Alb) | P43652 | AFM | Increase |
| Coagulation factor XIII B chain (Fibrin-stabilizing factor B subunit) (Protein-glutamine gamma-glutamyltransferase B chain) (Transglutaminase B chain) | P05160 | F13B | Increase |
| Complement C1q subcomponent subunit C | P02747 | C1QC | Decrease in male, steady in female |

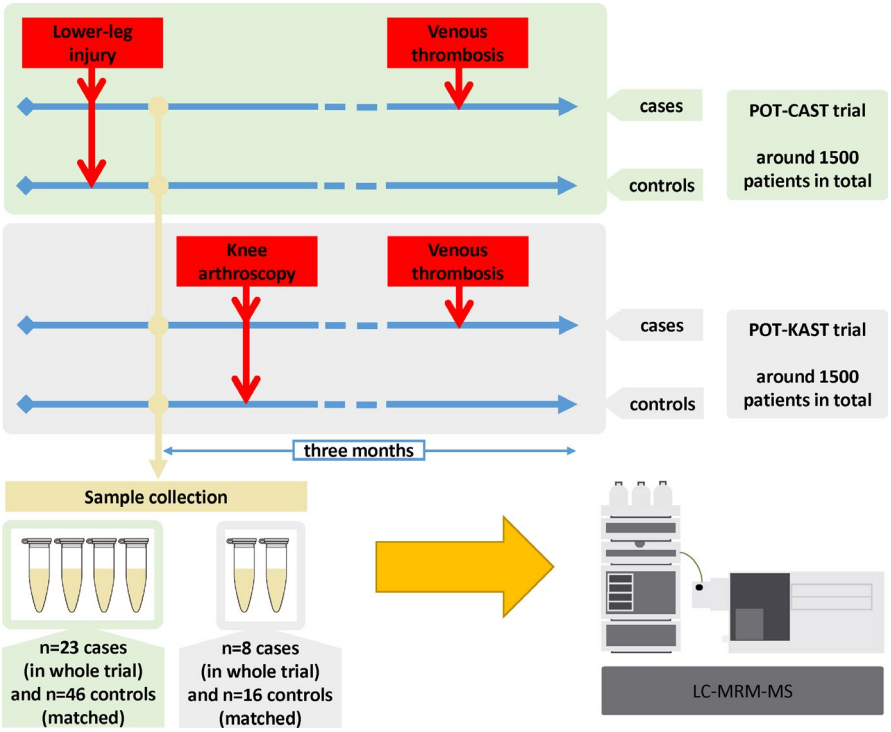



FIGURE 1 Overview of sample collection and measurement (the number of persons in the figure does not reflect the real number of patients in the trials)

exact gradient was as follows (time point in min, solution B %): 0 min, 2%; 2 min, 7%; 50 min 30%; 53min, 45%, 53.5 min, 80%; 55.5 min, 80%; 56 min, 2%. A post-column equilibration of 4 min was used after each sample analysis. The LC system was interfaced to a triple-quadrupole mass spectrometer (Agilent 6495/Agilent 6490) via a standard-flow ESI source, operated in the positive ion mode. The MRM acquisition parameters used for the quantitation were as follows: 3500 V capillary voltage, 300 V nozzle voltage, 11 L/min sheath gas flow at a temperature of 250°C, 15 L/min drying gas flow at a temperature of 150°C, 30 psi nebulizer gas pressure, 380 V fragmentor voltage, 5 V cell accelerator potential, and unit mass resolution in the first and third quadrupoles. For optimal peptide collision induced dissociation, peptide-specific collision energy values had previously been determined experimentally.

### 2.5 | Determination of protein concentration

The data was inspected using MassHunter Quantitative Analysis (version B.07.00; Agilent) and Skyline Version 3.1.[26] This involved peak inspection to ensure accurate selection, integration, and uniformity (of peak shape and retention time) of the measured peptides. For each peptide, the relative peak area ratios of the endogenous to the heavy labeled peptide were calculated. This ratio and the known concentration of internal standard were used to calculate the concentration of the endogenous peptide in the sample by comparison to a standard curve.[26] The criteria used for the standard curve regression analysis were $1/x^2$ regression weighting, <20% deviation in a given level's precision and accuracy for each concentration level.[26]

### 2.6 | Data and statistical analysis

We used Ward agglomerative method to perform the hierarchical clustering on the scaled and centered protein concentration values. Various methods for feature selection were used to determine top predictors including iterative random forest classification[29] for weighted selection, least absolute shrinkage and selection operator – LASSO,[30] grouped-lasso penalties – GAMSEL,[31] multiple logistic regression, as well as non-parametric Wilcoxon rank-sum test. We compared the performance of the different methods to conclude that non-parametric Wilcoxon rank-sum test outperformed the others. Comparison was based on assessing the strength of discrimination between cases and controls in a simple logistic regression discriminator. Logistic regression modeling was also used to calculate the area under the curve or C-statistics with the top predictors using cross validation with 80% and 20% of the data entries for training and testing, respectively. While we primarily used cross validation, parallel to that we also applied validation by bootstrapping for comparison. Here, instead of dividing the data entries into 80% and 20% training and testing sets, the training set is generated using sampling with replacement with the same number of entries as the whole data set, while the testing set is generated from the remaining entries that were not selected for training. For cross validation and bootstrapping, we used 100 repeats which were enough for the results to converge. When compared to 200 and 500 and 1000 repeats there were no significant changes in the reported AUC. We ensured that the training set included always at least one female and one male patient. We handled the two clinical situations separately in our analysis. Because sex-specific profiles based on plasma protein levels were previously reported in human and mouse,[32-36] we investigated whether stratification according to sex was necessary. For that, we identified top discriminating proteins between female and male subjects, including patients and controls, and we assessed the strength of the discrimination by C-statistics based on cross validation as described before. All data analysis steps and visualization were performed in R statistical software tool version 3.6.2. This study is exploratory, and the purpose of all feature selection methods was to reach an ordered list of variables from which the top discriminators are used in the final logistic regression model. These top discriminators are not affected by a false discovery rate correction. Rather than evaluating each protein individually, as one would do in a mechanistic study, we have evaluated the prediction models themselves using cross validation and bootstrapping and by providing the corresponding confidence intervals.

## 3 | RESULTS

We used a quantitative targeted proteomics panel that covers 270 proteins, which included the proteins that we were able to quantify in plasma in previous experiments.[22,23,27] We were able to detect 197 proteins and quantify 167 proteins in our samples (Table S2). For the quantification we have considered all protein concentrations within 50% and above of their lower limit of quantification. The measured proteins included many coagulation and complement factors, covering 64 proteins in the complement and coagulation cascade, of which 45 were successfully quantified. Figure 2 summarizes the determined concentrations from all 93 EDTA and 10 citrate plasma samples. In Supplementary Protein Concentration Report we included box and whisker plots for each protein individually along with the Wilcoxon rank-sum test results.

### 3.1 | Hierarchical clustering

In the hierarchical clustering in Figure 2, citrate plasma samples were grouped together. The clustering shows that almost all determined protein concentrations using citrate plasma are of lower levels compared to EDTA plasma samples. Median protein concentration ranged between 72% and 94% in citrate plasma of that of EDTA plasma. In addition, various proteins could only be measured in EDTA samples. This is mainly attributed to these samples containing a stronger anticoagulant that allows for less intrinsic protease and

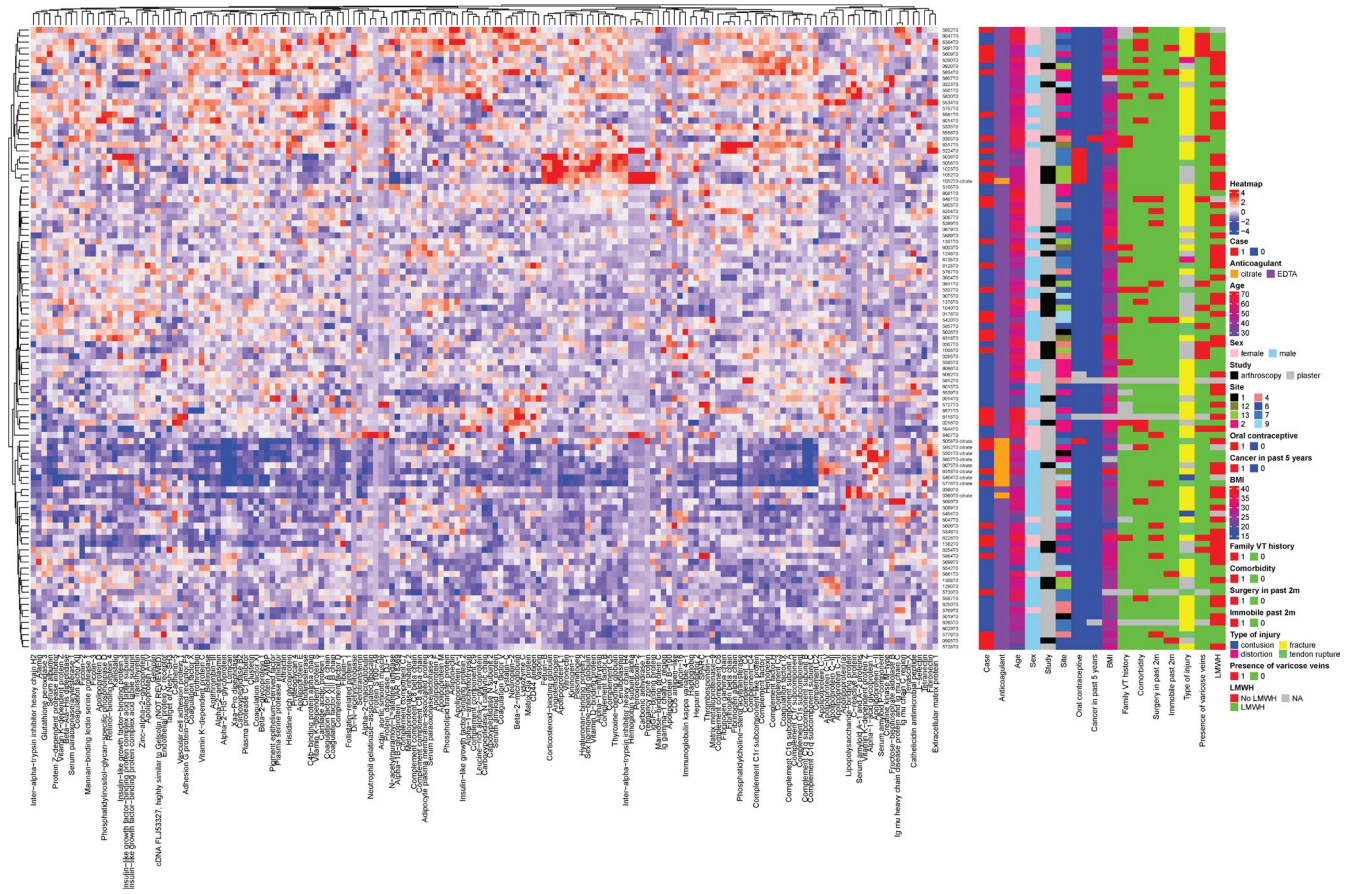



**FIGURE 2** Heatmap and hierarchical clustering of quantified proteins in all samples. The hierarchal clustering of patients is shown in the left side and of proteins on the top. The heatmap in the middle shows scaled protein concentrations (relative to mean of measured samples of each protein). Sample annotations are included on the right side along with the associated color legend. Annotations include case: with red for patients with VTE and blue for no VTE patients; anticoagulant used: citrate indicated in orange and EDTA in purple; patient age: on a color scale from 30 to 70 (blue to red); patient sex with pink refers to female and sky-blue to male; study: either knee arthroscopy in black or lower-leg injury with plaster cast in grey; site of sample collection (numbered); use of oral contraceptive with red indicating recorded use and blue for no use; cancer/malignancy in last 5 years with 1 indicating yes in red and 0 indicating no in blue; BMI: on a color scale from 15 to 40 (blue to red); family VT history indicating whether there were VT incidents in first degree relatives with red indicating yes, green no, and grey unknown; comorbidity: with red indicating yes, green no, and grey unknown; surgery as well as immobility in the past 2 months: both with red for yes, green for no, and grey for unknown; type of injury of lower-leg injury patients: with blue for contusion, yellow for fracture, dark pink for distortion, and green for tendon rupture; presence of varicose veins: with red indicating yes, green no, and grey unknown; and finally LMWH indicating wither LMWH was used in green, not used in red, while grey indicating missing values. The clustering shows that citrated plasma samples were grouped together. No grouping is associated with the sample collection site. A strong cluster constitutes female individuals who used oral contraceptives. Furthermore, samples were partially grouped according to sex

peptidase activities during the initial sample handling. Nonetheless, excellent correlations between protein levels in citrate and EDTA plasma samples were determined (Figure S1). Looking closely at Figure 2, we did not recognize any cluster associated with the sample collection site, which indicates that there was no, or negligible batch effect related to collection procedure.

A second strong cluster constituted female individuals who used oral contraceptives, which left a strong signature on plasma protein abundances leading to clear grouping of these subjects. Interestingly, this cluster showed a stronger grouping effect than treatment group (i.e., knee arthroscopy or cast immobilization). Furthermore, samples were partially grouped according to sex. Hence, sexual dimorphism protein profiles can be caught to some degree by the measured protein abundancies.

## 3.2 | Feature selection for best discrimination

After evaluating various strategies as mentioned in the Methods section, we concluded that the simplest strategy performed best. While random forest feature selection resulted in a protein set more suited for (unsupervised) hierarchical clustering, the feature selection based on Wilcoxon rank-sum test resulted in best discrimination using a regression model. In a simple yet robust approach, top

discriminating proteins were used in the predictors for each of the two clinical groups. As the number of discriminating proteins can vary; we studied the top 15 proteins and built our final regression models using a subset. Figure S2 shows the change in the C-statistics of the predictors for POT-CAST as well as POT-KAST based on consecutive addition of one discriminator at a time. Best discrimination was achieved using the selected discriminators as in Figure S2 and detailed below.

We investigated sexual dimorphism in our results, and a relatively good discrimination between male and female subjects, independently whether these were case or control subjects, was possible. Figure S3 shows a heatmap along with unsupervised hierarchical clustering based on the top 10 discriminators, which allowed a C-statistic of 94.3% (95% CI: 93.4–95.3%). Therefore, further analyses were stratified for sex.

### 3.3 | Venous thrombosis after knee arthroscopy and discrimination based on plasma protein profiles

Eight cases and 16 controls were included from the POT-KAST trial. This accounted for all cases present in the trial along with two matched controls per case. Figure S2-A represents the top 15 discriminating proteins as obtained by Wilcoxon rank-sum test along with the change in C-statistics as we include additional proteins. Figure 3 represents the six most discriminating proteins between cases and controls, stratified by sex, showing that cases had reduced levels of three apolipoproteins (C-II, C-III, C-IV), attractin, coagulation factor XIII, and prothrombin, i.e. APOC2, APOC3, APOC4, ATRN, F13B, and F2. Including these six proteins in a logistic regression model resulted in C-statistics of 88.1% (95%CI: 85.7–90.6%), calculated with repeated cross validation and similar values were reported with bootstrapping. Table 3 lists the regression coefficients and Figures S4 and S5 represent the hierarchal clustering obtained using these top discriminating proteins.

### 3.4 | Predicting venous thrombosis after lower-leg cast immobilization using plasma protein profiles

From the POT-CAST trial, 23 cases and 46 controls were included. Similar to POT-KAST, this accounted for all cases present in the trial ($n$ = 1519) along with two matched controls per case. Figure S2-B represents the top 15 discriminating proteins as obtained by Wilcoxon rank-sum test and illustrates how C-statistics change with the inclusion of top discriminating proteins. A logistic regression using the seven top discriminating proteins resulted in C-statistics of 79.6% (95% CI: 77.2–82.0%) estimated using cross validation, similar values were reported with bootstrapping. Figure 4 represents these seven proteins, showing different trends in cases versus controls in the abundance of apolipoprotein B-100, apolipoprotein E, alpha-2-antiplasmin, beta-2-microglobulin, afamin, coagulation factor XIII B chain, and complement C1q subcomponent subunit C, i.e., APOB, APOE, SERPINF2, B2 M, AFM, F13B, and C1QC. Table 3 lists the coefficients of the logistic regression model using these top discriminators. In comparison, using clinical predictors to discriminate between cases and controls such as age, sex, first-degree family history of VT event, comorbidity, presence of varicose veins, BMI, and previous surgery within 2 months before blood collection, all of which were part of various VT risk models such as L-TRiP,[16] the AUC was 63.2% (95% CI: 59.8–66.5%). We have also assessed the performance of discrimination when combining these clinical predictors with proteomics. Except for sex, the various possible combinations of the clinical predictors with the proteomic top discriminating proteins did not improve the classification performance over using protein abundancies alone. When using a binary sex predictor alongside the abundancies of the top seven discriminating proteins a slight improvement in C-statistic to from 79.6 to 81.0% (95% CI: 78.7–83.3%) was observed (Figure 4). Figures S6 and S7 represent the hierarchal clustering obtained using top discriminating proteins in lower-leg cast immobilization patients.

## 4 | DISCUSSION

Using quantitative targeted proteomics, we profiled 93 samples from the POT-(K)CAST trials. From approximately 3000 patients, all 31 patients who developed VT were included and matched with controls. Logistic regression models resulted in C-statistics of 88.1% and 79.6% for knee arthroscopy and lower leg cast immobilization groups, respectively.

This work relayed on two prospective studies, in which blood was sampled in patients before the thrombotic event (Figure 1). To that end, large cohorts are necessary to find sufficient number of cases, and this explains the low number of cases despite collecting more than 3000 samples. Ultimately, external validation is crucial to validate our results, nonetheless it is important to bear in mind that the current samples were collected in a multicentric effort at different sites, and that we used cross-validation as well as bootstrapping for our models. To reach stronger statistics than what we have in our work, for example 70 cases, a cohort of 10,000 patients would be necessary, which would possibly require multinational efforts making it quite challenging.

### 4.1 | Quantitative targeted mass spectrometry to measure protein abundances

Targeted protein quantification with internal standards is suited for longitudinal and multicentric studies because it references comparability to spiked-in internal standards. This was essential for the translational aspect of our initial question; namely can precise blood plasma protein abundancies be used to assess VT risk? We used targeted proteomics despite it not being widely available as a routine analysis technique in hospitals, which can be seen as a limiting factor

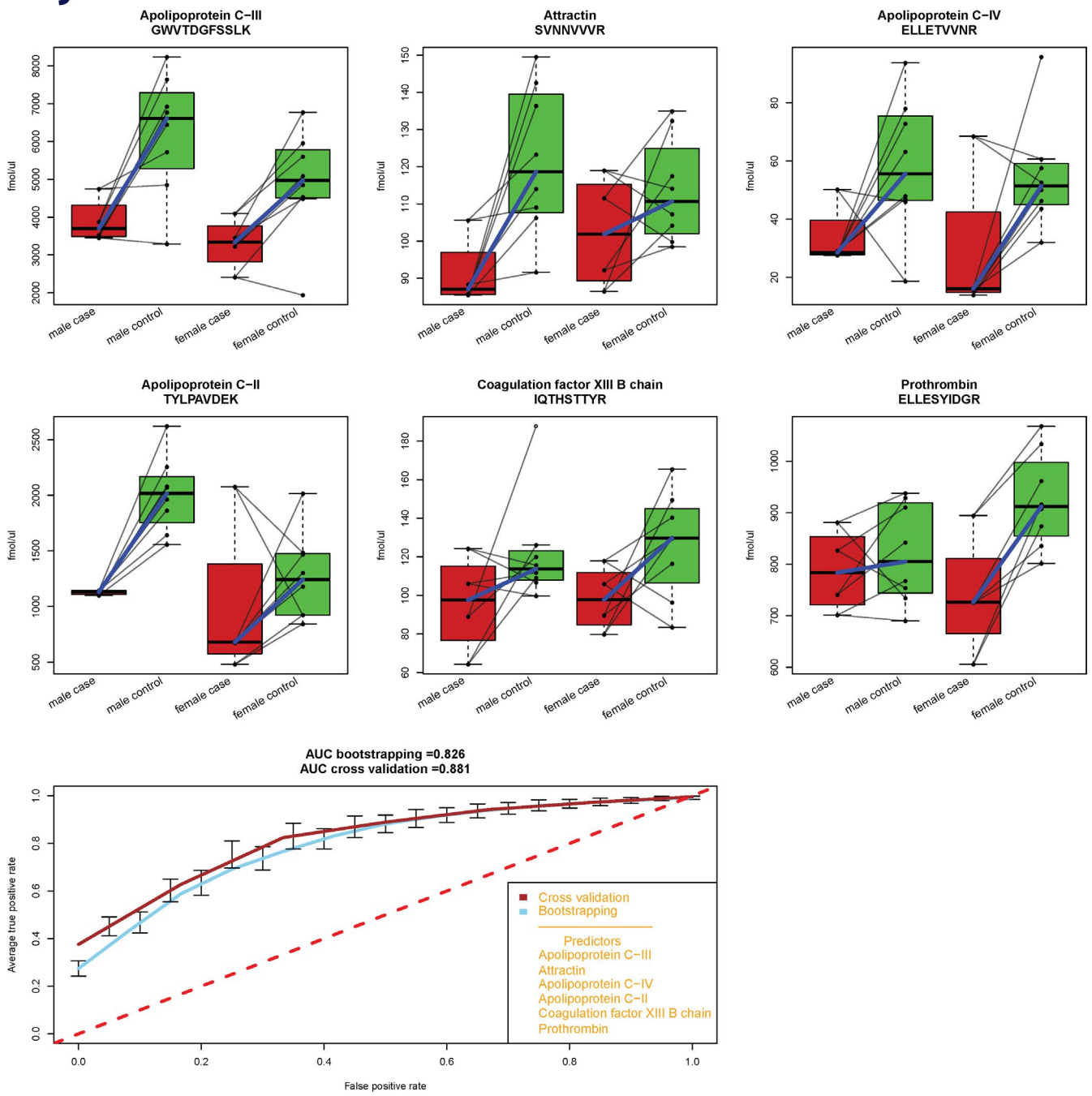



**FIGURE 3** Top discriminating proteins in POT-KAST and associated prediction model. Six discriminating proteins between cases and controls showed cases had reduced levels of APOC2, APOC3, APOC4, ATRN, F13B, and F2. A logistic regression model average AUC of 88.1% (95% CI: 85.7–90.6%, calculated with repeated cross validation with 100 repeats, and 80/20% training/testing sets)

for the generalizability of our work. Nonetheless, by identifying the few best proteins that are sufficient to predict risk, we envision that these can be used in a panel that does not necessary depend on mass spectrometry. We showed previously how targeted mass spectrometry-based proteomics results correlates well with antibody as well as activity assays.[37]

The proteomics community has demonstrated the ability to detect (not quantify) up to 912 proteins in plasma.[38] Here, we used a panel for 270 plasma proteins which were quantifiable in plasma in previous experiments.[22,23,27] Many assay development efforts, including ours, are performed under optimal conditions for sample collection and processing and therefore we do not expect to be able to quantify all proteins in real application. The panel used covered 64 proteins in the complement and coagulation cascade, of which 45 were successfully quantified (Figure 2). We were able to quantify many procoagulant and anticoagulant factors, as well as fibrinolysis proteins. This included coagulation factors II, V, IX, X, XI, XII, and XIII, as well as fibrinogen, prothrombin, thrombin-antithrombin complex,

**TABLE 3** Coefficients of the logistic regression for POT-KAST and POT-CAST

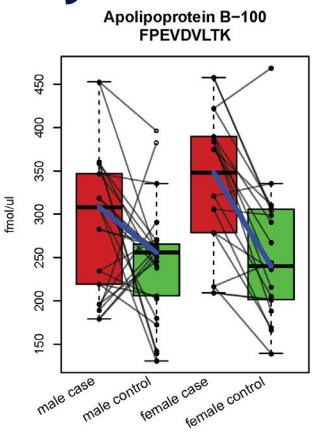
| | Cross validation | | | | Bootstrapping | | | |
|---|---|---|---|---|---|---|---|---|
| | | | 95% CI of mean | | | | 95% CI of mean | |
| | Median | Mean | Lower limit | Upper limit | Median | Mean | Lower limit | Upper limit |
| | Top discriminating proteins in POT-KAST and associated prediction model (Figure 3) | | | | | | | |
| (Intercept) | −25.828 | −27.895 | −29.491 | −26.299 | −18.462 | −19.753 | −21.483 | −18.023 |
| Apolipoprotein C III | −40.113 | −35.235 | −39.004 | −31.466 | −23.934 | −23.837 | −28.05 | −19.624 |
| Attractin | −41.372 | −43.984 | −46.625 | −41.342 | −30.096 | −30.274 | −32.902 | −27.647 |
| Apolipoprotein C IV | 6.278 | 0.967 | −2.451 | 4.386 | 1.308 | −2.024 | −5.617 | 1.57 |
| Apolipoprotein C II | −17.837 | −15.555 | −17.94 | −13.169 | −15.452 | −13.505 | −16.112 | −10.898 |
| Coagulation factor XIII B chain | −9.605 | −9.309 | −11.249 | −7.368 | −7.752 | −8.693 | −11.439 | −5.946 |
| Prothrombin | −22.532 | −20.426 | −22.767 | −18.085 | −18.984 | −15.224 | −18.436 | −12.012 |
| | Top discriminating proteins in POT-CAST and associated prediction models (Figure 4) | | | | | | | |
| (Intercept) | −1.279 | −1.332 | −1.367 | −1.298 | −1.316 | −1.461 | −1.533 | −1.389 |
| Apolipoprotein B 100 | 1.088 | 1.116 | 1.073 | 1.159 | 1.138 | 1.245 | 1.132 | 1.359 |
| Apolipoprotein E | −0.002 | −0.012 | −0.044 | 0.021 | −0.012 | 0.052 | −0.034 | 0.138 |
| Alpha 2 antiplasmin | 0.585 | 0.63 | 0.584 | 0.675 | 0.602 | 0.717 | 0.605 | 0.829 |
| Beta 2 microglobulin | 1.353 | 1.41 | 1.362 | 1.458 | 1.515 | 1.671 | 1.561 | 1.782 |
| Afamin | 0.276 | 0.277 | 0.237 | 0.317 | 0.35 | 0.349 | 0.257 | 0.442 |
| Coagulation factor XIII B chain | −0.381 | −0.282 | −0.341 | −0.222 | −0.391 | −0.305 | −0.458 | −0.152 |
| Complement C1q subcomponent subunit C | −1.549 | −1.605 | −1.65 | −1.56 | −1.698 | −1.812 | −1.907 | −1.718 |

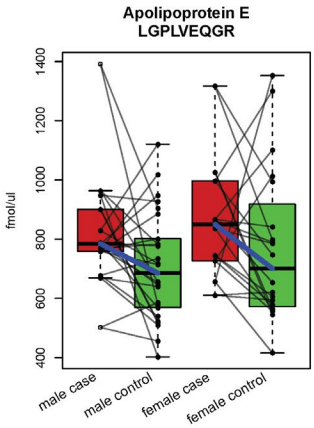
and von Willebrand factor.[39-45] Antithrombin, protein C, and protein S were also quantified.[46-48] Quantified fibrinolysis related proteins associated with VT risk included alpha2-antiplasmin, fibrinopeptide A, plasmin-alpha2-antiplasmin, plasminogen, tissue plasminogen activator, and Carboxypeptidase B2 (thrombin-activatable fibrinolysis inhibitor – TAFI).[49-52] However, three proteins associated with VT risk were detectable but not quantifiable, coagulation factor VII, thrombomodulin and p-selectin. Three additional proteins of interest, factor VIII, tissue factor pathway inhibitor, and plasminogen activator inhibitor-1 were not detected. Upon examination, we concluded that alternative surrogate peptides could improve quantifiability. Previously, we quantified coagulation factor VII using a different, longer proxy peptide; i.e. VAQVIIPSTYVPGTTNHDIALLR[37] versus VSQYIEWLQK in the current work. Longer sequences are harder to synthesize, however in this case the longer peptide is better. Another example is Factor VIII, which is a protein with high number of natural sequence variants – around 490 known and predicted. The large number of variants makes it hard to find suitable proteotypic peptides. The peptide used LHPTHYSIR contains few sites with various possible sequence variants; P2172L/Q/R, T2173A/I, H2174D, and R2178C/H/L. Although we always attempt to use peptides with no or insignificant (known) modifications and variances, our view on existing PTMs and variances is continuously updated with new published studies. Therefore, using multiple proteotypic peptides per protein is advantageous,[37] yet with a large number of targets, instrumentation speed is a limiting factor. Our panel required one-hour cycle per sample. Using longer gradient and adding additional fractionation steps allows measuring more targets, but increases cost.[53] Tissue factor pathway inhibitor can be quantified in a two-dimensional liquid chromatography method but not in a one-dimensional method.[53] In summary, balancing study objectives and number of multiplexed peptide assays is essential. For our objective, i.e. investigating predictability of VT risk using precise plasma protein concentrations, the panel used demonstrated the possibilities and future works may improve on our results.

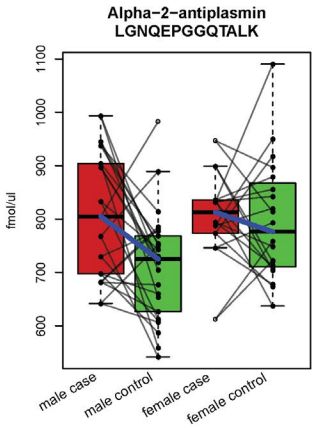
## 4.2 | Cross validation and bootstrapping

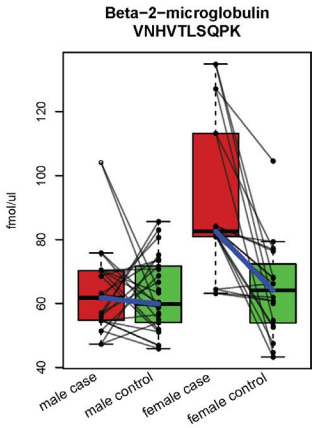
The POT-(K)CAST trials form a suitable testbed for developing a prediction model allowing two essential views of VT signature in blood, pre and post trauma. Cases from each cohort have their own control group and applying the same analytical and data analysis approaches allowed good comparability.

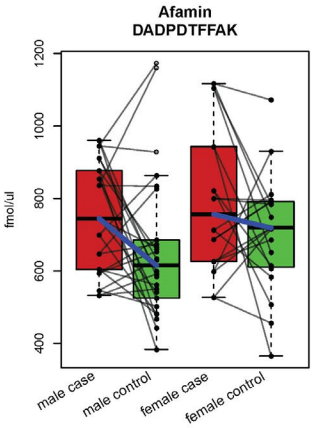
Validation in an independent cohort is important for testing prediction models. However, the low incidence rate in VT cases in general, makes it infeasible to replicate prospective trials like POT-(K)CAST. We opted therefore to perform computational approaches using cross validation and bootstrapping to assess predictability. Both approaches reported similar results as can be seen in Figures 3

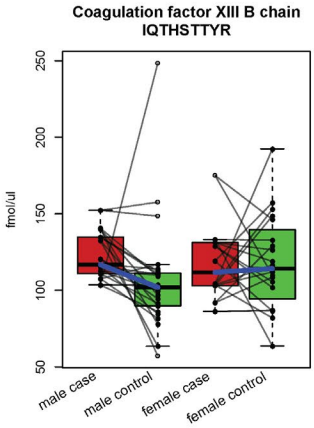

Apolipoprotein B−100
FPEVDVLTK
fmol/ul
male case
male control
female case
female control

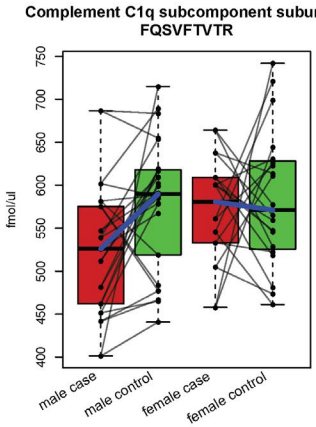

Apolipoprotein E
LGPLVEQGR
fmol/ul
male case
male control
female case
female control

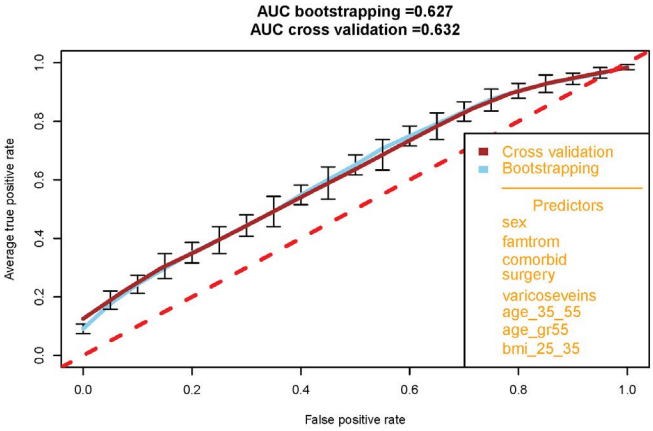

Alpha−2−antiplasmin
LGNQEPGGQTALK
fmol/ul
male case
male control
female case
female control

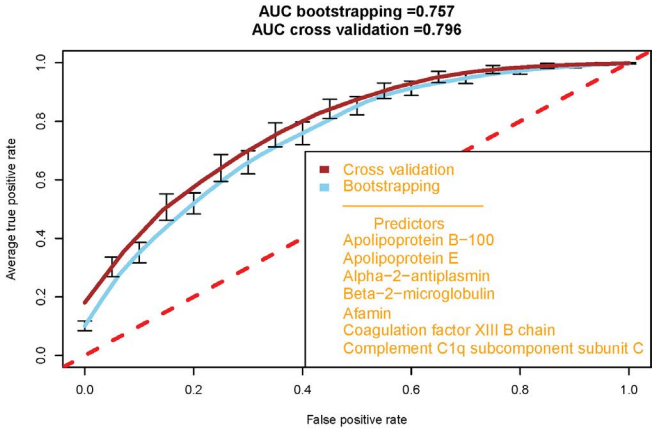

Beta−2−microglobulin
VNHVTLSQPK
fmol/ul
male case
male control
female case
female control


Afamin
DADPDTFFAK
fmol/ul
male case
male control
female case
female control


Coagulation factor XIII B chain
IQTHSTTYR
fmol/ul
male case
male control
female case
female control


Complement C1q subcomponent subunit (
FQSVFTVTR
fmol/ul
male case
male control
female case
female control


AUC bootstrapping =0.627
AUC cross validation =0.632
Average true positive rate
False positive rate
Cross validation
Bootstrapping
Predictors
sex
famtrom
comorbid
surgery
varicoseveins
age_35_55
age_gr55
bmi_25_35


AUC bootstrapping =0.757
AUC cross validation =0.796
Average true positive rate
False positive rate
Cross validation
Bootstrapping
Predictors
Apolipoprotein B−100
Apolipoprotein E
Alpha−2−antiplasmin
Beta−2−microglobulin
Afamin
Coagulation factor XIII B chain
Complement C1q subcomponent subunit C

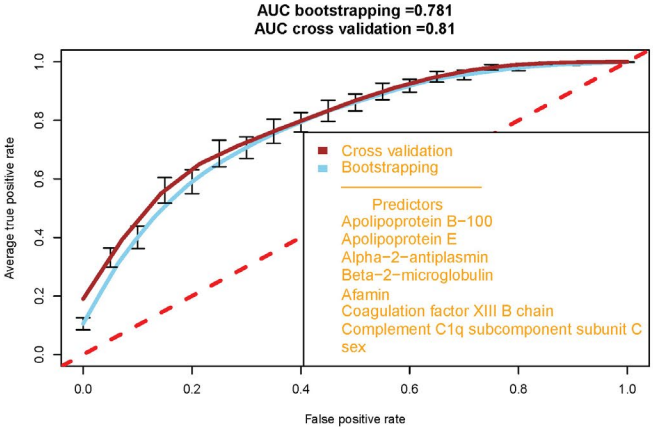

AUC bootstrapping =0.781
AUC cross validation =0.81
Average true positive rate
False positive rate
Cross validation
Bootstrapping
Predictors
Apolipoprotein B−100
Apolipoprotein E
Alpha−2−antiplasmin
Beta−2−microglobulin
Afamin
Coagulation factor XIII B chain
Complement C1q subcomponent subunit C
sex

**FIGURE 4** Top discriminating proteins in POT-CAST and associated prediction models. Seven proteins showed different trends in cases versus controls included APOB, APOE, SERPINF2, B2 M, AFM, F13B, and C1QC. A logistic regression these proteins resulted in an average AUC of 79.6% (95% CI: 77.2–82.0%, calculated with repeated cross validation with 100 repeats, and 80/20% training/testing sets). A logistic regression using clinical predictors (L-TRiP[16]) resulted in AUC of 63.2% (95% CI: 59.8–66.5%). Including sex with the top seven discriminating proteins resulted in a C-statistic of 81.0% (95% CI: 78.7–83.3%). In the three prediction models, the used predictors are mentioned in the legend. For the first prediction model we used clinical data alone, for the second protein abundances alone were used, and for the third we used protein abundance in addition of whether there was a surgery. famtrom: refer to history of VT in first degree relatives (Table 1)

and 4. Nonetheless, these computational approaches do not usually account for all variations in sample collection, storage, or other similar factors that are considered in an external validation using a separate cohort. However, because sample collection was multicentric, some of the aforementioned variabilities are inherited. The main goal of our work is to investigate whether, in principle, precise plasma protein profiling can be used to predict VT. To that end, cross validation and bootstrapping provided a good initial internal validation and further external validation as well as testing of each individual predictor should follow.

### 4.3 | Discriminating proteins in relation to VT after knee arthroscopy

Despite the low number of cases and controls in the POT-KAST we were able to pinpoint a discriminating profile based on the determined protein concentration (Figure 3). Having three apolipoproteins in the top discriminators indicates a relation between the levels of these proteins in blood plasma and the risk of VT. Apolipoproteins C-II, C-III, and C-IV are well characterized and are all constituent of circulating (triglyceride-rich) lipoproteins, are all relatively abundant in plasma, and are produced mainly in the liver, with C-II and C-IV being both part of the so-called APOE/C1/C4/C2 gene cluster[54] and CIII element of the APOA1/C3/A4/A5 gene cluster.[55] In both clusters, genes are coordinately expressed (i.e., share liver regulatory elements), and as such it is of interest that APOA4 (Figure 3) and APOE are part of the protein signature specific for patients that develop VT after arthroscopy or plaster cast, respectively (compare Figure 3 and Figure 4). As these C apolipoproteins are constituent of lipoproteins, their plasma levels are known to be associated with the risk of developing cardiovascular disease, in particular arterial disease.[56] While associations of C apolipoproteins with VT risk have not been reported in literature before, we recently found in case-control study that plasma C apolipoproteins (and also APOE) levels associate with the plasma levels of a number of pro-coagulant factors, suggesting a link between C apolipoproteins and VT.[57,58]

Prothrombin and factor XIIIb are both well-described coagulation factors, are also abundant in plasma, and are mainly produced in the liver. Prothrombin and Factor XIIIb are involved in fibrin network formation upon (vascular) injury, with factor XIIIb, in complex with Factor XIIIa, being activated by thrombin (i.e. the cleaved and activated form of prothrombin). Plasma prothrombin level is known to be associated with increased risk of VT.[59] Although less characterized, elevated Factor XIII (XIIIa and b in complex) levels are also associated with the risk of VT.[60,61] Recently, a study found a relation between factor XIIIb and sex, i.e. factor XIIIb levels were lower in male VT patients than in females.[62] As already mentioned above, we found previously that plasma C apolipoproteins associate with a panel of pro-coagulation factors and among these is prothrombin.[58] Of the six proteins that contributed to discrimination between cases and controls in knee arthroscopy patients, attractin is the most poorly characterized. Attractin has been reported to be involved in initial immune cell clustering during inflammatory response and to have a critical role in normal myelination in the central nervous system.[63] No link with vascular disease, including VT, has been described before, and therefore we can only speculate whether this protein relates to VT in a way beyond being solely a discriminating protein, i.e., a marker of something else.

### 4.4 | Discriminating proteins in relation to VT in patients with lower-leg cast immobilization

We measured an increase of APOB in cases compared to controls in the POT-CAST subjects. APOB has multiple functions in maintaining the homeostasis of liver-derived lipoproteins, especially of low-density lipoprotein (LDL). APOB is a component of LDL and ensures the binding of LDL to its receptor. In this way, it is able to regulate plasma levels of LDL and it has been identified to play a role in the pathophysiology of arterial thrombosis.[64-66] A direct association between APOB and VT is not known, but one previous study showed that APOB/LDL is able to bind von Willebrand Factor (vWF) and regulate the proteolytic cleavage of vWF by ADAMTS13. When vWF is cleaved, it becomes less procoagulant.[67] For beta-2-microglobulin, afamin, and complement C1q-subcomponent subunit C, that are all relatively abundant in plasma, literature does not provide clues on how these proteins may relate to VT. In contrast, alpha-2-antiplasmin is a well-documented inhibitor of plasmin, i.e. the enzyme that degrades fibrin networks as present in thrombosis. Of interest, alpha-2-antiplasmin may be cross-linked to fibrin by factor XIII.[68] Alpha-2-antiplasmin levels have been investigated for association with the risk of VT, though associations were modest (OR 1.2) in particular as compared to other proteins involved in regulation of fibrin degradation (PAI-1, tissue-type plasminogen activator – tPA, thrombin-activatable fibrinolysis inhibitor – TAFI[69]) and that are also part of the

MRM-assay used in the present study. As mentioned above, apolipoprotein E like the C apolipoproteins associates with VT in the MEGA case-control study, supporting some role for triglyceride-rich lipoprotein associated proteins in VT. Apolipoprotein E plays an important role in the distribution of vitamin K, which is necessary for the synthesis of coagulation factor II, VII, IX, X and protein C, S and Z.[70-72] Other studies state that certain genotypes of apoE are associated with higher risk of VT.[73]

### 4.5 | Prediction of VT using absolute quantitation of blood plasma proteins

Mass spectrometry-based proteomics is viewed largely as a high throughput method to elucidate and study disease and biological processes. We opted to answer the question whether we can use proteomics to predict VT, without in-depth exploration of mechanistic insights. Our results should not be interpreted in a mechanistic way, which requires a different type of modeling and analysis. With the exception of coagulation factor XIIIb, the protein discriminators in the knee arthroscopy cohort are different from patients with cast immobilization after a traumatic lower-leg injury. While this may warrant a future emphasis on studying factor XIIIb in the context of VT in orthopedic interventions, it is important to point out that blood collection in both groups took place at different time points in respect to trauma (Figure 1). Factor XIIIb showed increase in abundances when blood was drawn after trauma, i.e. POT-CAST, and its abundance was decreased when blood was drawn before trauma, i.e. POT-KAST, both compared to own matched controls. It is probable that in POT-CAST patients trauma-induced coagulopathy had started leading to increase of various procoagulants in the systemic circulation.[74] This is not the case in POT-KAST patients as those are at rest situation awaiting intervention, knee arthroscopy, which occurred post blood sampling. Previously, we studied the procoagulant state and found that patients with lower-leg injury had higher levels of procoagulant factors, while in knee arthroscopy patients these levels remained steady.[75] This suggests that quantitative proteomics has predictive value both in plasma reflecting a person's steady state, as in plasma in a person in whom a severe event has just taken place. Mechanistic associations between Factor XIIIb and the other discriminating proteins and the differences in plasma profile before and after trauma for VT cases is a matter of a future additional experiment.

We also examined all possible combinations of top discriminating proteins as predictors and the relationship to the prediction strength evaluated by the C-statistics. We evaluated a minimum C-statistics of 59.9% for POT-CAST and 75.6% for POT-KAST patients. A maximum C-statistics for POT-CAST patient prediction of 82.5% was obtained when APOB, SERPINF2, B2 M, and C1QC were considered. For prediction in the POT-KAST cohort, a maximum of 95.6% was achieved with APOC3, ATRN, and F2. When all proteins are included as predictors, the confidence intervals of the beta value of APOC4 for POT-KAST and of APOE for POT-CAST patients overlapped zero (Table 3). While these proteins are discriminating cases from control subjects in the corresponding cohort, and therefore are important to consider in our discussion, when combined with other top discriminating proteins their contribution to the regression model is reduced or eliminated. In summary, while optimizing the number of predictors is possible, the fact that the worst models still have predictive values supports our conclusion about the possibility to use precise protein concentration to predict VT.

## 5 | CONCLUSIONS

We presented initial results on the possibilities of predicting VT using accurate measurement of plasma protein concentrations by targeted quantitative proteomics. The results were demonstrated in samples from the POT-(K)CAST trials for assessing the risk of VT following traumatic lower-leg injury or knee arthroscopy. While these initial results, assessed by internal cross validation as well as bootstrapping, demonstrated the principle of prediction of VT using protein abundancies determined by targeted MRM proteomics, further external validation, especially due to the low sample size, is necessary. We showed how to measure 100+ proteins in low volumes of EDTA (<10 μl) and found a possible protein signature for VT. Protein concentration values from EDTA and citrate plasma showed very good correlation suggesting that our results in EDTA plasma is transferable to citrate plasma, both being the most common anticoagulant used in collecting plasma samples. Improving and extending the panel we used to quantify even more proteins is possible, yet focusing on the translational aspect is key. Ultimately, one does not want to measure hundreds of proteins to predict risk, rather quantifying only a few best predicting proteins should be sufficient. We followed this logic and showed that it is possible to perform prediction using up to seven proteins. Our results suggest quantitative targeted proteomics as a rapid promising low cost extension to current methods of blood testing.

### ACKNOWLEDGEMENTS

We would like to thank Ingeborg de Jonge and Petra Noordijk for the helpful support in sample collection and preparation. This work was supported by Genome Canada and Genome British Columbia's Genome Innovations Network grant (204PRO for operations; 214PRO for technology development), Technology Platform grant (264PRO), and Bioinformatics and Computational Biology grant (282PQP). C.H.B. is also grateful for financial support from the Segal McGill Chair in Molecular Oncology at McGill University (Montreal, Quebec, Canada), and for support from the Warren Y. Soper Charitable Trust and the Alvin Segal Family Foundation to the Jewish General Hospital (Montreal, Quebec, Canada). The study was also supported by a MegaGrant of the Ministry of Science and Higher Education of the Russian Federation (Agreement with Skolkovo Institute of Science and Technology, No. 075-10-2019-083).

## CONFLICT OF INTEREST

C.H.B. is the Chief Scientific Officer of MRM Proteomics, Inc., the co-founder and Chief Technology Officer of Creative Molecules, Inc. and Chief Technology Officer of Molecular You. The other authors declare no competing financial interests.

## AUTHOR CONTRIBUTIONS

S.C.C. designed the current study and set up the original clinical trials. C.E.T., B.N. and R.A.A. collected the samples. Y.M. and C.H.B. performed the proteomics analysis and collected the data. Y.M. performed the data analysis, generated the figures, and wrote the initial draft of the manuscript. Y.M., C.E.T., B.N., R.A.v.A, CHB, F.R.R., B.v.V. and S.C.C. interpreted the results and wrote the manuscript. All authors read and contributed to the final version of the text.

## ORCID

*Yassene Mohammed* https://orcid.org/0000-0003-3265-3332
*Carolina E. Touw* https://orcid.org/0000-0002-7993-5209
*Banne Nemeth* https://orcid.org/0000-0002-1214-5923
*Bart J. van Vlijmen* https://orcid.org/0000-0003-0208-6453
*Suzanne C. Cannegieter* https://orcid.org/0000-0003-4707-2303

## REFERENCES

1. Naess IA, Christiansen SC, Romundstad P, Cannegieter SC, Rosendaal FR, Hammerstrom J. Incidence and mortality of venous thrombosis: a population-based study. *J Thromb Haemost*. 2007;5:692-699.
2. Piazza G, Goldhaber SZ. Chronic thromboembolic pulmonary hypertension. *N Eng J Med*. 2011;364:351-360.
3. Baldwin MJ, Moore HM, Rudarakanchana N, Gohel M, Davies AH. Post-thrombotic syndrome: a clinical review. *J Thromb Haemost*. 2013;11:795-805.
4. Lijfering WM, Rosendaal FR, Cannegieter SC. Risk factors for venous thrombosis - current understanding from an epidemiological point of view. *Br J Haematol*. 2010;149:824-833.
5. Rosendaal FR. Venous thrombosis: a multicausal disease. *Lancet*. 1999;353:1167-1173.
6. van Adrichem RA, Debeij J, Nelissen RG, Schipper IB, Rosendaal FR, Cannegieter SC. Below-knee cast immobilization and the risk of venous thrombosis: results from a large population-based case-control study. *J Thromb Haemost*. 2014;12:1461-1469.
7. van Adrichem RA, Nelissen RG, Schipper IB, Rosendaal FR, Cannegieter SC. Risk of venous thrombosis after arthroscopy of the knee: results from a large population-based case-control study. *J Thromb Haemost*. 2015;13:1441-1448.
8. van Adrichem RA, Nemeth B, Algra A, et al. Thromboprophylaxis after knee arthroscopy and lower-leg casting. *New Engl J Med*. 2017;376:515-525.
9. Hippisley-Cox J, Coupland C. Development and validation of risk prediction algorithm (QThrombosis) to estimate future risk of venous thromboembolism: prospective cohort study. *BMJ*. 2011;343:d4656.
10. Sultan AA, West J, Grainge MJ, et al. Development and validation of risk prediction model for venous thromboembolism in postpartum women: multinational cohort study. *BMJ*. 2016;355:i6253.
11. Timp JF, Braekkan SK, Lijfering WM, et al. Prediction of recurrent venous thrombosis in all patients with a first venous thrombotic event: The Leiden Thrombosis Recurrence Risk Prediction model (L-TRRiP). *PLoS Med*. 2019;16:e1002883.
12. Nemeth B, Douillet D, le Cessie S, et al. Clinical risk assessment model to predict venous thromboembolism risk after immobilization for lower-limb trauma. *EClinicalMedicine*. 2020;20:100270.
13. Barbar S, Noventa F, Rossetto V, et al. A risk assessment model for the identification of hospitalized medical patients at risk for venous thromboembolism: the Padua prediction score. *J Thromb Haemost*. 2010;8:2450-2457.
14. Kucher N, Koo S, Quiroz R, et al. Electronic alerts to prevent venous thromboembolism among hospitalized patients. *N Engl J Med*. 2005;352:969-977.
15. Nemeth B, van Adrichem RA, van Hylckama VA, et al. Venous thrombosis risk after arthroscopy of the knee: derivation and validation of the L-TRiP(ascopy) score. *Thromb Haemost*. 2018;118:1823-1831.
16. Nemeth B, van Adrichem RA, van Hylckama VA, et al. Venous thrombosis risk after cast immobilization of the lower extremity: derivation and validation of a clinical prediction score, L-TRiP(cast), in three population-based case-control studies. *PLoS Med*. 2015;12:e1001899.
17. Douillet D, Nemeth B, Penaloza A, et al. Venous thromboembolism risk stratification for patients with lower limb trauma and cast or brace immobilization. *PLoS One*. 2019;14:e0217748.
18. Doerr A. Mass spectrometry-based targeted proteomics. *Nat Methods*. 2013;10:23.
19. Picotti P, Bodenmiller B, Aebersold R. Proteomics meets the scientific method. *Nat Methods*. 2013;10:24-27.
20. Gillette MA, Carr SA. Quantitative analysis of peptides and proteins in biomedicine by targeted mass spectrometry. *Nat Methods*. 2013;10:28-34.
21. Carr SA, Abbatiello SE, Ackermann BL, et al. Targeted peptide measurements in biology and medicine: best practices for mass spectrometry-based assay development using a fit-for-purpose approach. *Mol Cell Proteomics*. 2014;13:907-917.
22. Mohammed Y, Kootte RS, Kopatz WF, et al. The intestinal microbiome potentially affects thrombin generation in human subjects. *J Thromb Haemost*. 2020;18(3):642-650.
23. Gaither C, Popp R, Mohammed Y, Borchers CH. Determination of the concentration range for 267 proteins from 21 lots of commercial human plasma using highly multiplexed multiple reaction monitoring mass spectrometry. *Analyst*. 2020;145(10):3634-3644.
24. Bhowmick P, Roome S, Borchers CH, Goodlett DR, Mohammed Y. An update on MRMAssayDB: a comprehensive resource for targeted proteomics assays in the community. *J Proteome Res*. 2021;20:2105-2115.
25. Mohammed Y, Domanski D, Jackson AM, et al. PeptidePicker: a scientific workflow with web interface for selecting appropriate peptides for targeted proteomics experiments. *J Proteomics*. 2014;106:151-161.
26. Mohammed Y, Pan J, Zhang S, Han J, Borchers CH. ExSTA: External standard addition method for accurate high-throughput quantitation in targeted proteomics experiments. *Proteomics Clin Appl*. 2018;12(2):1600180.
27. Percy AJ, Mohammed Y, Yang J, Borchers CH. A standardized kit for automated quantitative assessment of candidate protein biomarkers in human plasma. *Bioanalysis*. 2015;7:2991-3004.
28. Percy AJ, Chambers AG, Yang J, Hardie DB, Borchers CH. Advances in multiplexed MRM-based protein biomarker quantitation toward clinical utility. *Biochim Biophys Acta*. 2014;1844:917-926.
29. Tin KH. Random decision forests. Proceedings of 3rd International Conference on Document Analysis and Recognition, 1995, 278-82 vol.1.
30. Tibshirani R. Regression shrinkage and selection via the Lasso. *J Roy Stat Soc B Met*. 1996;58:267-288.
31. Simon N, Tibshirani R. Standardization and the group lasso penalty. *Stat Sin*. 2012;22:983-1001.
32. Rietjens I, Vervoort J, Maslowska-Gornicz A, Van den Brink N, Beekmann K. Use of proteomics to detect sex-related differences

in effects of toxicants: implications for using proteomics in toxicology. *Crit Rev Toxicol*. 2018;48:666-681.
33. Curran AM, Fogarty Draper C, Scott-Boyer MP, et al. Sexual dimorphism, age, and fat mass are key phenotypic drivers of proteomic signatures. *J Proteome Res*. 2017;16:4122-4133.
34. Choi JW, Aseer KR, Chaudhari HN, Mukherjee R, Choi M, Yun JW. Gender dimorphism in regulation of plasma proteins in streptozotocin-induced diabetic rats. *Proteomics*. 2013;13:2482-2494.
35. Saad MF, Damani S, Gingerich RL, et al. Sexual dimorphism in plasma leptin concentration. *J Clin Endocrinol Metab*. 1997;82:579-584.
36. Tilburg J, Michaud SA, Maracle CX, et al. Plasma protein signatures of a murine venous thrombosis model and Slc44a2 knockout mice using quantitative-targeted proteomics. *Thromb Haemost*. 2020;120(03):423-436.
37. Mohammed Y, van Vlijmen BJ, Yang J, et al. Multiplexed targeted proteomic assay to assess coagulation factor concentrations and thrombosis-associated cancer. *Blood Adv*. 2017;1:1080-1087.
38. Geyer PE, Voytik E, Treit PV, et al. Plasma proteome profiling to detect and avoid sample-related biases in biomarker studies. *EMBO Mol Med*. 2019;11:e10427.
39. Rietveld IM, Lijfering WM, le Cessie S, et al. High levels of coagulation factors and venous thrombosis risk: strongest association for factor VIII and von Willebrand factor. *J Thromb Haemost*. 2019;17:99-109.
40. Rietveld IM, Bos MHA, Lijfering WM, et al. Factor V levels and risk of venous thrombosis: The MEGA case-control study. *Res Pract Thromb Haemost*. 2018;2:320-326.
41. van Hylckama VA, van der Linden IK, Bertina RM, Rosendaal FR. High levels of factor IX increase the risk of venous thrombosis. *Blood*. 2000;95:3678-3682.
42. de Visser MC, Poort SR, Vos HL, Rosendaal FR, Bertina RM. Factor X levels, polymorphisms in the promoter region of factor X, and the risk of venous thrombosis. *Thromb Haemost*. 2001;85:1011-1017.
43. Meijers JC, Tekelenburg WL, Bouma BN, Bertina RM, Rosendaal FR. High levels of coagulation factor XI as a risk factor for venous thrombosis. *New Engl J Med*. 2000;342:696-701.
44. Ota S, Wada H, Abe Y, et al. Elevated levels of prothrombin fragment 1 + 2 indicate high risk of thrombosis. *Clin Appl Thromb Hemost*. 2008;14:279-285.
45. Ginsberg JS, Brill-Edwards P, Panju A, et al. Pre-operative plasma levels of thrombin-antithrombin III complexes correlate with the development of venous thrombosis after major hip or knee surgery. *Thromb Haemost*. 1995;74:602-605.
46. Croles FN, Borjas-Howard J, Nasserinejad K, Leebeek FWG, Meijer K. Risk of venous thrombosis in antithrombin deficiency: a systematic review and bayesian meta-analysis. *Semin Thromb Hemost*. 2018;44:315-326.
47. Pintao MC, Ribeiro DD, Bezemer ID, et al. Protein S levels and the risk of venous thrombosis: results from the MEGA case-control study. *Blood*. 2013;122:3210-3219.
48. Dahm A, Van Hylckama VA, Bendz B, Rosendaal F, Bertina RM, Sandset PM. Low levels of tissue factor pathway inhibitor (TFPI) increase the risk of venous thrombosis. *Blood*. 2003;101:4387-4392.
49. Meltzer ME, Doggen CJ, de Groot PG, Rosendaal FR, Lisman T. The impact of the fibrinolytic system on the risk of venous and arterial thrombosis. *Semin Thromb Hemost*. 2009;35:468-477.
50. Meltzer ME, Lisman T, de Groot PG, et al. Venous thrombosis risk associated with plasma hypofibrinolysis is explained by elevated plasma levels of TAFI and PAI-1. *Blood*. 2010;116:113-121.
51. Folsom AR, Cushman M, Heckbert SR, Rosamond WD, Aleksic N. Prospective study of fibrinolytic markers and venous thromboembolism. *J Clin Epidemiol*. 2003;56:598-603.
52. Meltzer ME, Doggen CJ, de Groot PG, Rosendaal FR, Lisman T. Fibrinolysis and the risk of venous and arterial thrombosis. *Curr Opin Hematol*. 2007;14:242-248.
53. Percy AJ, Simon R, Chambers AG, Borchers CH. Enhanced sensitivity and multiplexing with 2D LC/MRM-MS and labeled standards for deeper and more comprehensive protein quantitation. *J Proteomics*. 2014;106:113-124.
54. Ordovas JM, Schaefer EJ. Genetic determinants of plasma lipid response to dietary intervention: the role of the APOA1/C3/A4 gene cluster and the APOE gene. *Br J Nutr*. 2000;83(Suppl 1):S127-S136.
55. Lai CQ, Parnell LD, Ordovas JM. The APOA1/C3/A4/A5 gene cluster, lipid metabolism and cardiovascular disease risk. *Curr Opin Lipidol*. 2005;16:153-166.
56. Olivieri O, Martinelli N, Girelli D, et al. Apolipoprotein C-III predicts cardiovascular mortality in severe coronary artery disease and is associated with an enhanced plasma thrombin generation. *J Thromb Haemost*. 2010;8:463-471.
57. Morelli VM, Lijfering WM, Bos MHA, Rosendaal FR, Cannegieter SC. Lipid levels and risk of venous thrombosis: results from the MEGA-study. *Eur J Epidemiol*. 2017;32:669-681.
58. Orsi FA, Lijfering WM, Van der Laarse A, et al. Association of apolipoproteins C-I, C-II, C-III and E with coagulation markers and venous thromboembolism risk. *Clin Epidemiol*. 2019;11:625-633.
59. Chinthammitr Y, Vos HL, Rosendaal FR, Doggen CJ. The association of prothrombin A19911G polymorphism with plasma prothrombin activity and venous thrombosis: results of the MEGA study, a large population-based case-control study. *J Thromb Haemost*. 2006;4:2587-2592.
60. Kucher N, Schroeder V, Kohler HP. Role of blood coagulation factor XIII in patients with acute pulmonary embolism. Correlation of factor XIII antigen levels with pulmonary occlusion rate, fibrinogen, D-dimer, and clot firmness. *Thromb Haemost*. 2003;90:434-438.
61. Van Hylckama VA, Komanasin N, Ariens RA, et al. Factor XIII Val34Leu polymorphism, factor XIII antigen levels and activity and the risk of deep venous thrombosis. *Br J Haematol*. 2002;119:169-175.
62. Mezei ZA, Katona E, Kallai J, et al. Factor XIII levels and factor XIII B subunit polymorphisms in patients with venous thromboembolism. *Thromb Res*. 2017;158:93-97.
63. Duke-Cohan JS, Gu J, McLaughlin DF, Xu Y, Freeman GJ, Schlossman SF. Attractin (DPPT-L), a member of the CUB family of cell adhesion and guidance proteins, is secreted by activated human T lymphocytes and modulates immune cell interactions. *Proc Natl Acad Sci USA*. 1998;95:11336-11341.
64. Bayly GR. CHAPTER 37 - Lipids and disorders of lipoprotein metabolism. In: Marshall WJ, Lapsley M, Day AP, Ayling RM, eds. *Clinical Biochemistry: Metabolic and Clinical Aspects* (Third Edition). London: Churchill Livingstone;2014:702-736.
65. Vrablík M, Ceska R, Horínek A. Major apolipoprotein B-100 mutations in lipoprotein metabolism and atherosclerosis. *Physiol Res*. 2001;50:337-343.
66. van den Berg MJ, van der Graaf Y, de Borst GJ, Kappelle LJ, Nathoe HM, Visseren FLJ. Low-density lipoprotein cholesterol, non-high-density lipoprotein cholesterol, triglycerides, and apolipoprotein b and cardiovascular risk in patients with manifest arterial disease. *Am J Cardiol*. 2016;118:804-810.
67. Cao W, Abdelgawwad MS, Li J, Zheng XL. Apolipoprotein B100/Low-density lipoprotein regulates proteolysis and functions of von willebrand factor under arterial shear. *Thromb Haemost*. 2019;119:1933-1946.
68. Fraser SR, Booth NA, Mutch NJ. The antifibrinolytic function of factor XIII is exclusively expressed through alpha(2)-antiplasmin cross-linking. *Blood*. 2011;117:6371-6374.
69. Meltzer ME, Doggen CJM, de Groot PG, Rosendaal FR, Lisman T. Plasma levels of fibrinolytic proteins and the risk of myocardial infarction in men. *Blood*. 2010;116:529-536.
70. Kohnke H, Sorlin K, Granath G, Wadelius M. Warfarin dose related to apolipoprotein E (APOE) genotype. *Eur J Clin Pharmacol*. 2005;61:381-388.
71. Rost S, Fregin A, Koch D, Compes M, Muller CR, Oldenburg J. Compound heterozygous mutations in the gamma-glutamyl carboxylase gene cause combined deficiency of all vitamin K-dependent blood coagulation factors. *Br J Haematol*. 2004;126:546-549.

72. Sadler JE. Medicine: K is for koagulation. *Nature*. 2004;427:493-494.
73. Nagato LC, de Souza Pinhel MA, de Godoy JM, Souza DR. Association of ApoE genetic polymorphisms with proximal deep venous thrombosis. *J Thromb Thrombolysis*. 2012;33:116-119.
74. Hayakawa M. Pathophysiology of trauma-induced coagulopathy: disseminated intravascular coagulation with the fibrinolytic phenotype. *J Intensive Care*. 2017;5:14.
75. Touw CE, Nemeth B, Schipper IB, Nelissen RGHH & Cannegieter SC Lower-leg injury leads to a procoagulant state [Abstract]. Abstracts of the 64th Annual Meeting of the Scientific Standardization Committee of the International Society on Thrombosis and Haemostasis. Dublin, 2018 July 18-21, 323. Abstract no 668.

## SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

**How to cite this article:** Mohammed Y, Touw CE, Nemeth B, et al. Targeted proteomics for evaluating risk of venous thrombosis following traumatic lower-leg injury or knee arthroscopy. *J Thromb Haemost*. 2022;20:684–699. doi:10.1111/jth.15623