

**Proteomics
Clinical Applications**

**Multiplexed LC/ESI-MRM-MS-based Assay for Identification
of Coronary Artery Disease Biomarkers in Human Plasma**

Journal: *Clinical Applications*

Manuscript ID prca.201700111.R1

Wiley - Manuscript type: Research Article

Date Submitted by the Author: 16-Aug-2018

Complete List of Authors: Anwar, Mohammad; Molecular You Corporation
Dai, Liying; NCE CECR PROOF Centre of Excellence
Wilson-McManus, Janet; NCE CECR PROOF Centre of Excellence
Smith, Derek; University of Victoria - Genome BC Proteomics Centre
Francis, Gordon; University of British Columbia, Department of Medicine;
Providence Heart and Lung Institute and the James Hogg Research Center,
St. Paul's Hospital
Borchers, Christoph; University of Victoria - Genome BC Proteomics
Centre, ; University of Victoria, Biochemistry & Microbiology
McManus, Bruce; University of British Columbia, Department of Pathology
and Laboratory Medicine; NCE CECR PROOF Centre of Excellence
Hill, John; University of British Columbia, Department of Pathology and
Laboratory Medicine; Providence Heart and Lung Institute and the James
Hogg Research Center, St. Paul's Hospital
Cohen-Freue, Gabriela; University of British Columbia, Department of
Statistics

Keywords: proteomic classifier, Coronary artery disease

Multiplexed LC/ESI-MRM-MS-based Assay for Identification of Coronary Artery Disease Biomarkers in Human Plasma

Mohammad A Anwar^{*1}, Darlene Liying Dai^{*2}, Janet Wilson-McManus², Derek Smith³, Gordon A. Francis^{4,5}, Christoph H Borchers^{3,6}, Bruce M. McManus^{2,7}, John S. Hill^{3,7}, Gabriela V. Cohen Freue^{2,8}

¹Molecular You Corporation, Vancouver, British Columbia, Canada; ²NCE CECR PROOF Centre of Excellence, Vancouver, British Columbia, Canada; ³Genome BC Proteomics Centre, University of Victoria, Victoria, British Columbia, Canada; ⁴Providence Heart and Lung Institute and the James Hogg Research Center, St. Paul's Hospital, University of British Columbia, Vancouver, British Columbia, Canada; ⁵Department of Medicine, University of British Columbia, Vancouver, BC, Canada; ⁶Department of Biochemistry and Microbiology, University of Victoria, Victoria, British Columbia, Canada; ⁷Department of Pathology and Laboratory Medicine, University of British Columbia, Vancouver, BC, Canada; ⁸Department of Statistics, University of British Columbia, Vancouver, British Columbia, Canada. *equal contributors

Correspondence should be addressed to: Dr. Gabriela V. Cohen Freue, Assistant Professor, ESB 3146-2329 West Mall, University of British Columbia, Vancouver, BC V6T 1Z4. Tel: 604-822-3710 Email: gcohen@stat.ubc.ca

A2AP	Alpha 2-antiplasmin	FV	Coagulation factor V
ApoCII	Apolipoprotein C II	FX	Coagulation factor X
ApoL1	Apolipoprotein L1	LDL	Low density lipoprotein
C3	Complement C3	MAC	Membrane attack complex
CAD	Coronary artery disease	PON1	Paraoxonase 1
CD5L	CD5 antigen-like	SHBG	Sex hormone binding globulin
CIQA	Complement C1q subcomponent subunit A	vWF	von Willebrand factor
CRP	C-reactive protein		

Keywords: Proteomics, Biomarkers, Coronary artery disease, MRM-MS, Classification

Total number of words: 9568

Statement of clinical relevance

Coronary artery disease (CAD) is the most common form of heart disease and occurs when arteries in the heart become blocked. If left untreated, CAD can lead to other serious problems such as angina, heart attack, stroke, and death. Currently, CAD is one of the most common causes of mortality in the developed world and has an estimated economic burden of more than \$170 billion per year and increasing. Therefore, accurate diagnosis, rapid testing, and risk management strategies is critical in reducing the morbidity, mortality, and socioeconomic burden of this disease. Although many potential circulating biomarkers have been elucidated and aid in the diagnosis of CAD pathogenesis, their rapid and reproducible measurement is required before they can be considered for routine use in risk evaluation for CAD. Our study illustrates the potential of using an accurate and reproducible LC/ESI-MRM-MS assay to identify promising biomarkers of CAD and may be used as a robust and accurate clinical tool for diagnosis and risk stratification.

Abstract

Purpose: A highly-multiplexed LC/ESI-MRM-MS based assay was developed for accurate and reproducible identification of coronary artery disease (CAD) biomarkers in human plasma.

Experimental design: The assay was used to measure 242 stable isotope labeled peptide standards and native peptides from 100 putative biomarkers of cardiovascular diseases in tryptic digests of plasma from subjects with (n=70) and without (n=45) angiographic evidence of CAD and no subsequent cardiovascular mortality during follow-up. *Results:* Extensive computational and statistical analysis revealed 16 plasma proteins associated with CAD, namely alpha 2-antiplasmin, apolipoprotein CII, apolipoprotein L1, C reactive protein, CD5 antigen-like, coagulation factor V, coagulation factor X, complement C3, complement C1q subcomponent subunit A, fibronectin, inter alpha trypsin inhibitor heavy chain H1, paraoxonase, protein S, sex hormone binding globulin, transferrin, and von Willebrand factor. The identified proteins were combined into a logistic regression with high classification performance (cross-validated AUC=0.75). When combined with a separate score computed from markers currently used in the clinic with similar performance, the AUC increased to 0.84. Similar results were observed in an independent set of subjects (n=87). *Conclusions and clinical relevance:* If externally validated, the assay and identified biomarkers can improve CAD risk stratification.

Introduction

Coronary artery disease (CAD) is one of the most common causes of death for both men and women in the developed world, responsible for approximately 1 in every 4 deaths [1]. CAD is estimated to affect about 17 million people in the United States and 1.6 million people in Canada [1-3]. The estimated direct and indirect economic burden of CAD in North America is more than \$170 billion per year [1]. The significant morbidity, mortality, and socioeconomic importance of this disease make timely accurate diagnosis and cost-effective prevention and management of CAD of the utmost importance.

CAD pathogenesis occurs when plaque builds up on the inner arterial walls [4] blocking the flow of oxygenated blood through the arteries to the heart muscle. This partial obstruction can lead to chest pain (angina) or a heart attack. Most heart attacks happen when a blood clot associated with plaque suddenly cuts off the hearts' blood supply, causing permanent heart damage. Over time, without intervention, CAD can also weaken the heart muscle and contribute to heart failure [4]. Chronic stable angina, the initial diagnostic manifestation of CAD, is only evident in approximately 50% of all subjects [5]. Other cases of CAD remain asymptomatic or present anginal “equivalents” symptoms, such as epigastric discomfort, dyspnea, fatigue, or faintness. Coronary angiography must be performed as quickly as possible in high-risk subjects with CAD. To date, there is still an unmet clinical need to enhance our current risk stratification based mainly on classic risk factors of CAD including family history, age, hyperlipidemia, hypertension, smoking, obesity and diabetes, and to determine who among an at-risk population has developed clinical CAD.

1
2
3 Although many potential circulating biomarkers may aid in diagnosis of CAD, their rapid
4 and reproducible measurement is required before they can be considered for routine use in risk
5 evaluation for CAD [6, 7]. Recent developments in quantitative proteomics technologies have
6 enabled the identification of potential biomarkers of various diseases and conditions [8, 9]. In
7 particular, liquid chromatography-electrospray ionization-multiple reaction monitoring mass
8 spectrometry (LC/ESI-MRM-MS) is a highly sensitive and reproducible method for the targeted
9 quantitation of protein/peptide abundance over a wide dynamic range of protein concentrations in
10 complex biological samples [10]. Since many of the potential CAD proteomic biomarkers are
11 circulating plasma proteins, LC/ESI-MRM-MS results in a reliable multiplex proteomics
12 platform for this discovery study. Although LC/ESI-MRM-MS is commonly used for the
13 quantification of small molecules in pharmaceutical, clinical, and environmental applications
14 [10-12], its use in discovery studies remains limited because of the requirement that the peptide
15 targets be known in advance.
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33

34 We used a large set of 100 potential protein biomarkers of cardiovascular diseases to
35 develop an accurate and reproducible discovery multiplexed LC/ESI-MRM-MS assay. Protein
36 levels were then measured in 202 plasma samples from a cohort of subjects who underwent
37 coronary angiography for suspected CAD. Based on our LC/ESI-MRM-MS platform results and
38 data analysis, we identified a panel of 16 plasma proteins and combined them into a logistic
39 regression score with high classification performance (cross-validated AUC=0.75, and a
40 sensitivity and specificity of 74% and 55%, respectively on an independent test set). A second
41 classifier was built from markers routinely used in the clinic, which showed similar performance
42 in our cohort of subjects (cross-validated AUC=0.76, test sensitivity and specificity of 77% and
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 36%, respectively). Interestingly, a classifier based on the average of both classifiers'
4 probabilities (ensemble classifier [13]) increased the cross-validated AUC to 0.84.
5
6

7
8 The results of this study are a solid first step toward the use of multiplexed LC/ESI-
9 MRM-MS assays in proteomic discovery studies, and for the development of a robust and
10 accurate tool to determine who, among an at-risk population, has developed clinical CAD.
11
12
13
14

15 16 **Materials and Methods**

17 18 19 *Study Population and Study Design*

20
21
22 The subjects for this study were derived from the Vancouver Coronary Angiography Cohort,
23 which has previously been described in detail [14-16]. All subjects provided signed consent
24 forms. This study was approved by the Research Ethics Board of St. Paul's Hospital, Vancouver,
25 Canada. Briefly, the cohort was comprised of men and women who were referred to two major
26 teaching hospitals in Vancouver for selective coronary angiography between 1992 and 1995.
27 Clinical indications for angiography included stable angina, previous myocardial infarction,
28 aortic valve disease, and mitral valve regurgitation. Subjects with unstable angina or myocardial
29 infarction within the preceding 2 months were excluded from the study. CAD presence was
30 defined by the presence of any lesion causing $\geq 20\%$ (diameter reduction) stenosis, while severe
31 CAD was defined by the presence of any lesion causing $\geq 50\%$ stenosis, as semi-quantitatively
32 visualized in major myocardial blood vessels (rounded to the nearest 10% stenosis). The subjects
33 were followed passively until November 2007. During the follow-up period, deaths related to a
34 cardiovascular disease were determined using the World Health Organization International
35 Classification of Disease-10th revision mortality codes I20-I25 and I60-I69.
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53

54
55
56 A case-control design was used to identify a panel of proteomic markers of CAD by
57
58
59

1
2
3 comparing plasma samples from 70 subjects who had angiographic evidence of CAD (case,
4 CAD+) with samples from 45 subjects who had no angiographic evidence of CAD (control,
5 CAD-). To control for confounding factors, samples from the case and the control groups in the
6 discovery set were approximately matched according to age and gender, and otherwise randomly
7 selected from the subset of subjects who did not die from a cardiovascular disease within the
8 follow-up period. A set of 87 additional subjects who were not included in the discovery cohort
9 was used as an independent test set for initial validation of the results of this study. Most of the
10 samples in this set correspond to subjects with angiographic evidence of CAD, who died from a
11 cardiovascular disease within the follow-up period (CAD+CVD+, n=73). The other 14 samples
12 in the test set correspond to CAD+ (n=3) and CAD- subjects (n=11) who did not die within the
13 follow-up period (CVD-) but could not be matched with samples in the discovery cohort.
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28

29 *Clinical Data and Plasma Sample Collection*

30
31
32 Information on demographics, history of cardiovascular disease, diabetes, hypertension,
33 smoking, and renal insufficiency in the patient and the patient's family was obtained from a
34 detailed questionnaire administrated to every patient by a nurse or cardiologist at the time of the
35 angiography. Body mass index (BMI) was calculated from each subject's weight and height.
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
Waist circumference and blood pressure were also measured.

Before angiography, a fasting blood sample was collected from a vein in the antecubital fossa in EDTA (ethylenediamine tetra-acetate) Vacutainer tubes (Becton Dickinson, Franklin Lakes, New Jersey) [14-16]. Plasma lipid levels and apolipoprotein B levels were immediately determined as previously described [14-16]. Blood samples were centrifuged immediately and aliquoted into smaller vials and frozen in -70°C until analysis. Plasma samples for the LC/ESI-

1
2
3 MRM-MS analysis were sent in 2011 to the University of Victoria (UVic) Genome BC
4
5 Proteomics Centre (Victoria, BC, Canada).
6
7

8 9 *Synthesis and Purification of Isotopically Labeled Standard Peptides*

10
11 All reagents were American Chemical Safety (ACS) grade or higher. All solvents used, including
12 water, were LC/MS grade. For this study, a total of 242 targeted peptides corresponding to 100
13 unique proteins (mostly involved in immunity, inflammation, coagulation, and cardiovascular
14 disease) were selected from the relevant literature (Table A1 in the Supplementary Material).
15
16 Isotopically labeled peptide standards were synthesized as previously described [17]. Briefly,
17 Fmoc (N-(9-fluorenyl) methoxycarbonyl) chemistry was performed on a Protein Technologies
18 Prelude peptide synthesizer (Tucson, AZ), using isotopically labeled amino acids, [$^{13}\text{C}_6$, $^{15}\text{N}_2$]- or
19 [$^{13}\text{C}_6$]-Lys (98% isotopic enrichment) and [$^{13}\text{C}_6$, $^{15}\text{N}_4$]- or [$^{13}\text{C}_6$]-Arg (98% isotopic enrichment)
20 which were purchased from Cambridge Isotope Laboratories (Andover, MA). The synthetic
21 peptides were dried and precipitated with ether. Peptides were purified by reversed-phase HPLC
22 (Ultimate 3000, Dionex, Thermo Scientific, Sunnyvale, CA) while monitoring the peptide elution
23 at 230 nm. Fractions containing > 80% of the target peptide by MALDI-TOF analysis were
24 combined and lyophilized. The peptides were then acid hydrolyzed, and the purity was
25 determined by amino acid analysis Capillary Zone Electrophoresis (CZE) using a P/ACE™
26 MDQ System (Beckman Coulter, Fullerton, CA) [17]. The absolute concentrations of the
27 synthetic peptides, as determined by amino acid analysis, were adjusted by the percent purity of
28 each synthetic peptide as determined by CZE. This corrects for any possible contribution from
29 incorrect or partial synthesis products.
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Preparation of Plasma Tryptic Digests and Addition of Internal Standard Peptides

All sample manipulations including in-solution digestion, stable-isotope-labeled standard (SIS) peptide addition, and solid phase extraction (SPE) cleanup steps were performed using a Tecan Freedom Evo150 liquid-handling robot (Männedorf, Switzerland). In each case, plasma samples were thawed and kept at 4°C. A single “standard plasma” sample was created by mixing six 20- μL samples from each group. This standard sample was added to each plate for the generation of a 6-point standard curve covering a 30-fold concentration range. Samples were mixed by vortexing for 10 sec, and centrifuged at 4°C for 5 min at 2,000 \times g. The samples were diluted 1/10 by transferring 8 μL of each sample into 72 μL of 25 mM ammonium bicarbonate in a polypropylene PCR microtitre plate and kept at 4°C, and 30 μL of each diluted plasma sample was denatured and diluted by transferring it to a deep well (1.6 mL) microtitre plate containing 30 μL of 10% w/v sodium deoxycholate (NaDOC) and 174.5 μL of 25 mM ammonium bicarbonate. Disulphide bonds were reduced by the addition of 26.1 μL of 50 mM tris(2-carboxyethyl)phosphine (TCEP, in 25 mM ammonium bicarbonate) (5 mM final concentration) and incubating at 60°C for 30 min in a dry-air incubator. Free sulfhydryl groups were alkylated by the addition of 29.0 μL of 100 mM iodoacetamide (in 25 mM ammonium bicarbonate) (10 mM final concentration) and incubating at 37°C for 30 min in a dry-air incubator. The remaining iodoacetamide was quenched by addition of 29.0 μL of 100 mM DTT (in 25 mM ammonium bicarbonate) (10 mM final concentration) and incubating at 37°C for 30 min in a dry-air incubator. 10.5 μL of TPCK-treated trypsin (0.8 $\mu\text{g}/\mu\text{L}$, Worthington, in 25 mM ammonium bicarbonate) was added to each sample and the final volume of each digest was 329 μL , and digestion was conducted at 37°C for 16 hours in a dry air incubator.

1
2
3 Digestion was quenched by the addition of an acidified SIS peptide mixture (in formic
4 acid) to give a final formic acid concentration of 0.5% v/v. This lowers the pH to <3, which
5
6 inactivates trypsin, and precipitates sodium deoxycholate. Samples were then centrifuged for 10
7
8 min at $2000 \times g$ (23°C) to pellet the sodium deoxycholate (NaDOC) precipitate. The sample
9
10 supernatant was desalted and concentrated by solid phase extraction using a Waters Oasis HLB
11
12 $\mu\text{Elution}$ plate (2 mg) for the 85 sample analyses or Waters Oasis HLB 1 cc (10 mg) cartridges
13
14 (Waters, Milford, MA). Briefly, the resin was rinsed with methanol and equilibrated with water.
15
16 The sample was loaded and washed with water, and eluted in 50% ACN (0.1% formic acid).
17
18 Eluted samples were frozen and lyophilized to dryness overnight. Prior to LC/ESI-MRM-MS
19
20 analysis, samples were rehydrated in a volume of 0.1% v/v formic acid (mobile phase A)
21
22 corresponding to a 1/70 dilution of plasma ($\square 1 \mu\text{g}/\mu\text{L}$, based on an initial plasma-protein
23
24 concentration of 70 mg/mL).
25
26
27
28
29

30 31 *LC/ESI-MRM/MS Analysis of Plasma Digests*

32
33
34 An Agilent 1290 Infinity UHPLC system was used to directly inject 10 μL of desalted plasma
35
36 digest samples ($\square 10 \mu\text{g}$, representing 143 nL of raw plasma) onto a reversed phase analytical
37
38 column (150 mm \times 2.1 mm i.d., Agilent ZORBAX Eclipse Plus C18 Rapid Resolution HD
39
40 column, 1.8 μm particle size) that was maintained at a column temperature of 50°C . Samples
41
42 were then separated using a flow rate of 400 $\mu\text{L}/\text{min}$ with a 2-min linear gradient from 3 to 13%
43
44 solvent B, then a 15-min linear gradient from 13 to 19% solvent B (90% v/v acetonitrile, 0.1%
45
46 v/v formic acid), followed by a 6-min linear gradient from 19 to 27% solvent B, then a 3-min
47
48 linear gradient from 27 to 43% solvent B, and finally a 2-min linear gradient from 43 to 90%
49
50 solvent B. The analytical column is then reconditioned by holding solvent B at 90% for 4 min
51
52 prior to ramping back down to 3% solvent B, and re-equilibrating the column for 4 min with 3%
53
54
55
56
57
58
59
60

1
2
3 solvent B. A blank, solvent injection (20 min analyses at 500 $\mu\text{L}/\text{min}$) was analyzed between all
4
5 samples to reduce sample carryover on the HPLC column.
6
7

8
9 An Agilent 6490 Triple Quadrupole MS with a high-flow iFunnel technology ionization
10
11 source, controlled by Agilent's MassHunter Workstation software (version B.04.01), was used for
12
13 all LC/ESI-MRM-MS sample analyses. All acquisition methods used the following parameters:
14
15 3500 V capillary voltage, a sheath gas flow of 11 L/min. (UHP nitrogen), a 200°C sheath gas
16
17 temperature, an MS operating pressure of 5.08×10^{-5} Torr, and Q1 and Q3 set to unit resolution.
18
19

20
21 MRM acquisition methods were constructed using 3 ion pairs per peptide with fragment
22
23 ion specific tuned CE voltages and retention time constraints. A default 380 V fragmentor
24
25 voltage and 5 V cell accelerator potential were used for all MRM ion pairs, and the dynamic
26
27 MRM option was used for all data acquisition with a target cycle time of 1 second and a 0.9 min
28
29 MRM detection window.
30
31

32 33 *LC/ESI-MRM-MS Data Analysis*

34
35

36 All LC/ESI-MRM-MS data was processed using Agilent MassHunter Quantitative Analysis
37
38 software (Agilent B.04.00) with the Agilent Integrator algorithm for peak integration set with
39
40 default values. All integrated peaks were manually inspected to ensure correct peak detection and
41
42 accurate integration. Linear regression of all calibration curves was performed using a standard
43
44 $1/x^2$ (x = concentration) weighting option to aid in covering a wide dynamic range [18]. All
45
46 peptides were initially targeted using 3 MRM ion pairs per peptide to verify retention times and
47
48 ensure absence of any signal interference. The multiple ion-pair signals for each peptide were
49
50 compared to detect the presence of interferences by ensuring that the relative signal intensity of
51
52 ion pairs was consistent between the heavy and natural forms of all peptides. Calculated
53
54
55
56
57
58
59
60

1
2
3 concentrations were derived from the quantifier MRM transition, with two qualifier transitions
4 acting to verify retention times and reveal any signal interference. Qualifier ion pairs were used
5 to detect the presence of interferences by ensuring that the relative signal intensity of ion pairs
6 was consistent between the heavy and endogenous forms of all peptides. We have aimed for a
7 minimum of 10 points across a peak and each peptide was measured with 3 transitions. Qualifier
8 and quantifier transitions are reported as a Supplementary File.
9
10
11
12
13
14
15
16
17

18 The quality of the linear regression curve is a good indicator of the quality of the data for
19 the specific target. The accuracy of each standard point was calculated based on the determined
20 linear regression curve and should be between 80-120%. Low signal standard points, such as
21 Std. A, are more likely to have accuracies outside this range and were sometimes excluded from
22 the linear regression curve. Most standard points should be within the 80-120% accuracy range
23 for a reliable determination of target concentration.
24
25
26
27
28
29
30
31

32 Relative target responses are reported as Area Ratio (RR: relative response), which is the
33 ratio of the integrated area of the endogenous (natural) peak to the integrated area of the
34 corresponding standard (SIS) peptide. Relative responses were used in subsequent statistical
35 analyses to ease the comparison of samples processed in different batches.
36
37
38
39
40
41

42 *Statistical Analyses*

43
44

45 We performed an extensive quality control of the LC/ESI-MRM-MS data as previously
46 described [19]. To control the quality of our data, we have examined the accuracies of the
47 standards points and the limit of quantification of the endogenous targets. Peptides with overall
48 low accuracy and/or values below the limit of quantitation or not measured in more than 50% of
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 the samples were eliminated from this analysis (see Table A1c in the Supplementary Material).
4
5 The relative responses were log-transformed (base 2).
6
7

8
9 For any protein containing missing values (i.e., peptides undetected in certain samples), if
10 another peptide is available and measured for that protein, then its value is used to impute the
11 missing value. Otherwise, half of the minimum value observed throughout the assay is used to
12 impute. This approach reduces the repetition of a single value in multiple samples introduced by
13 imputation. This approach reduces the repetition of a single value in multiple samples introduced by
14 imputation. This approach reduces the repetition of a single value in multiple samples introduced by
15 imputation. This approach reduces the repetition of a single value in multiple samples introduced by
16 imputation. This approach reduces the repetition of a single value in multiple samples introduced by
17 imputation. This approach reduces the repetition of a single value in multiple samples introduced by
18 imputation. This approach reduces the repetition of a single value in multiple samples introduced by
19 imputation. This approach reduces the repetition of a single value in multiple samples introduced by
20 imputation. This approach reduces the repetition of a single value in multiple samples introduced by

21 For proteins measured with multiple peptides, the reported protein level was based on the
22 peptide with the highest intensity in the majority of the samples analyzed. Results in this study
23 are robust to using the average level among all available peptides (see Supplementary Material).
24 Proteins with very small variation across the majority of the samples in the discovery set were
25 not analyzed. Results of the quality control process and the list of selected peptides for each
26 protein are shown in Table A1 in the Supplementary Material. All remaining protein values were
27 standardized to have a mean level of 0 and standard deviation of 1 for subsequent statistical
28 analyses. Results of the quality control process and the list of selected peptides for each
29 protein are shown in Table A1 in the Supplementary Material. All remaining protein values were
30 standardized to have a mean level of 0 and standard deviation of 1 for subsequent statistical
31 analyses. Results of the quality control process and the list of selected peptides for each
32 protein are shown in Table A1 in the Supplementary Material. All remaining protein values were
33 standardized to have a mean level of 0 and standard deviation of 1 for subsequent statistical
34 analyses. Results of the quality control process and the list of selected peptides for each
35 protein are shown in Table A1 in the Supplementary Material. All remaining protein values were
36 standardized to have a mean level of 0 and standard deviation of 1 for subsequent statistical
37 analyses. Results of the quality control process and the list of selected peptides for each
38 protein are shown in Table A1 in the Supplementary Material. All remaining protein values were
39 standardized to have a mean level of 0 and standard deviation of 1 for subsequent statistical
40 analyses. Results of the quality control process and the list of selected peptides for each
41 protein are shown in Table A1 in the Supplementary Material. All remaining protein values were
42 standardized to have a mean level of 0 and standard deviation of 1 for subsequent statistical
43 analyses. Results of the quality control process and the list of selected peptides for each
44 protein are shown in Table A1 in the Supplementary Material. All remaining protein values were
45 standardized to have a mean level of 0 and standard deviation of 1 for subsequent statistical
46 analyses. Results of the quality control process and the list of selected peptides for each
47 protein are shown in Table A1 in the Supplementary Material. All remaining protein values were
48 standardized to have a mean level of 0 and standard deviation of 1 for subsequent statistical
49 analyses. Results of the quality control process and the list of selected peptides for each
50 protein are shown in Table A1 in the Supplementary Material. All remaining protein values were
51 standardized to have a mean level of 0 and standard deviation of 1 for subsequent statistical
52 analyses. Results of the quality control process and the list of selected peptides for each
53 protein are shown in Table A1 in the Supplementary Material. All remaining protein values were
54 standardized to have a mean level of 0 and standard deviation of 1 for subsequent statistical
55 analyses. Results of the quality control process and the list of selected peptides for each
56 protein are shown in Table A1 in the Supplementary Material. All remaining protein values were
57 standardized to have a mean level of 0 and standard deviation of 1 for subsequent statistical
58 analyses. Results of the quality control process and the list of selected peptides for each
59 protein are shown in Table A1 in the Supplementary Material. All remaining protein values were
60 standardized to have a mean level of 0 and standard deviation of 1 for subsequent statistical

40 Differences in protein levels between CAD(+) and CAD(-) groups of the remaining
41 proteins were tested using a Student's *t*-test. Protein levels of significant proteins (un-adjusted *p*-
42 values < 0.05) were combined into a protein classifier probability score using logistic regression.
43 The logistic regression model is a linear combination of the explanatory variables (e.g., protein
44 levels) to explain and predict the probability of positive CAD. Similarly, traditional clinical
45 markers (Table 1) were used to build a logistic regression and predict the probability of positive
46 CAD. While the statistical significance of each marker is relevant to identify potential markers, it
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 is important to note that it may not affect the prediction performance when all markers are
4 combined in the logistic model (see Shmueli G. (2010) for a full discussion on this topic [20]). In
5
6 addition, the combination of several markers can achieve a performance superior to that of
7
8 individual features (see Figure A2 in the Supplementary Material). Since our study is focused on
9
10 the *joint* contribution of proteins and clinical features to predict CAD classes, we have not
11
12 adjusted the p-values of the univariate tests for multiple comparison testing. Both classifiers were
13
14 combined taking the average of the proteomics-MRM and the clinical estimated probabilities to
15
16 build a third classifier (ensemble classifier) [13].
17
18
19
20
21
22

23 In the absence of an external test cohort and considering the special characteristics of the
24
25 87 samples left aside as a test set, the three classifiers were first evaluated using 100 runs of 5-
26
27 fold stratified cross-validations. In each cross-validation run, the case and the control groups
28
29 were randomly split into 5 sets (a.k.a folds). One fold from each group was left out to form a test
30
31 set. The remaining samples were combined into a training set to build the classifiers and predict
32
33 the probability that the samples left out correspond to subjects with CAD. The test set that was
34
35 left out was not used to identify the proteins in the biomarker panel or build the logistic
36
37 regression classifiers. The predicted probabilities of each classifier were pooled to estimate each
38
39 area under the receiver operating curve (AUC). Similarly, we estimated the sensitivity and
40
41 specificity of the classifiers using a cut-off probability equal to 0.5. The receiver operating
42
43 characteristic curve (ROC) was used to illustrate these measures at other cut-off points. Since the
44
45 performance of the classifiers depends on the random assignment of samples into training and
46
47 test sets, the process was repeated 100 times and the performance measures obtained from all
48
49 runs were averaged for each classifier. The statistical analysis in this study was performed using
50
51
52
53
54
55
56
57
58
59
60

1
2
3 the software R (www.r-project.org) and Bioconductor (www.bioconductor.org). The AUC and
4
5 ROC were estimated using the ROCR [21] and cvAUC [22] packages .
6
7

8 9 **Results**

10 11 *Cohort Clinical Characteristics*

12
13
14
15 A set of 70 case subjects with angiographic evidence of CAD (CAD+), and 45 control subjects
16
17 without angiographic evidence of CAD (CAD-) were selected from the Vancouver Coronary
18
19 Angiography Cohort (Table 1) for proteomic-based biomarker discovery using the LC/ESI-
20
21 MRM-MS assay developed for this study. Some of the traditional clinical markers of CAD,
22
23 including body mass index (BMI), waist circumference, high-density lipoprotein cholesterol
24
25 (HDL-C), ratio of total cholesterol to HDL-C, previous incidence of diabetes, and previous
26
27 history of cardiovascular problems were significantly different between our CAD+ and CAD-
28
29 groups (Table 1). Since samples in both groups were approximately matched according to age
30
31 (61.96 ± 3.21 and 61.47 ± 1.41 years of age for the CAD- and the CAD+ groups, respectively, *t*-
32
33 test *p*-value equal to 0.3) and gender (80% and 84% of male in the CAD- and CAD+ groups,
34
35 respectively, Pearson's chi-squared test *p*-value equal to 0.7), these variables were not included in
36
37 the analysis of traditional clinical risk factors. Similarly, the majority of subjects in this study
38
39 were Caucasian so we did not analyze the association of ethnicity with CAD. The prevalence of
40
41 diabetes is higher in the CAD group. Thus, we also used the classifiers developed in this study to
42
43 classify only those samples without diabetes (n=97 from the discovery set) to address any
44
45 potential confounding effect. Results demonstrate that the performances of the three classifiers
46
47 developed in this study remain almost unchanged when evaluated on this subset of patients
48
49 without diabetes (see the Supplementary Material for more details).
50
51
52
53
54
55
56
57
58
59
60

LC/ESI-MRM-MS Assay for Biomarkers Discovery

We developed a proteomic LC/ESI-MRM-MS assay to measure a comprehensive list of putative CAD peptides on these samples. Table A1 in the Supplementary Material lists all the putative proteins and their corresponding peptide sequence(s) measured with the tailored LC/ESI-MRM-MS assay designed in this study. A detailed description of part of the assay has been previously given by our group [23]. Among the 242 measured peptides, we analyzed a total of 122 peptides, corresponding to 66 proteins, which passed the quality controls and were present in more than 50% of the discovery samples (see Table A1 in the Supplementary Material). The correlation between peptides levels from the same protein varies in the set of analyzed proteins with a median correlation of 0.94. The peptide with highest intensity in the majority of the analyzed samples was used to estimate the levels of proteins measured with multiple peptides (identified in Table A1 in the Supplementary Material).

A total of 16 proteins had a biologically and statistically significant differential abundance between subjects with and without CAD (t -test un-adjusted p -value < 0.05 , Table 2), and were used to build a logistic regression classifier. Figure 1 illustrates the protein levels of this discovery panel. Alpha 2 antiplasmin (A2AP), apolipoprotein CII (ApoCII), apolipoprotein L1 (ApoL1), C reactive protein (CRP), CD5 antigen-like (CD5L), coagulation factor V (F5), coagulation factor X (F10), complement C3 (C3), complement C1q subcomponent subunit A (C1QA), inter alpha trypsin inhibitor heavy chain H1 (ITIH1), paraoxonase (PON1), protein S (PROTS), and transferrin (TF) are, on average, more abundant in CAD⁺ subjects compared to CAD⁻ subjects. Fibronectin (FN1), von Willebrand factor (vWF) and sex hormone binding globulin (SHBG) are, on average, less abundant in CAD⁺ subjects compared to CAD⁻ subjects.

1
2
3 Figure 2 shows the pairwise correlations of all peptide levels from these 16 proteins in the
4 discovery panel. A cluster analysis of these proteins is given in Figure A3 of the Supplementary
5 discovery panel. A cluster analysis of these proteins is given in Figure A3 of the Supplementary
6 Material. As expected, multiple peptides from the same protein tend to be highly correlated
7 compared to peptides from other proteins (dark blue blocks across the diagonal line). Although
8 for some proteins the correlation between some peptides is low (e.g., the correlation between
9 SQHLDNFSNQIGK and LSEGASYLDHTFPAEK of coagulation factor V is 0.28, represented by
10 a light blue in Figure 2), in general, all proteins have at least one pair of highly correlated
11 peptides. The median of the minimum pairwise correlation between peptides of the same protein
12 is 0.74; the median of the maximum pairwise correlation between peptides of the same protein is
13 0.77. Not surprisingly, we also observed a number of large correlations among distinct proteins,
14 illustrating potential biological relationships among proteins in the human plasma proteome.
15
16
17
18
19
20
21
22
23
24
25
26
27
28

29 The 16 identified proteins were combined into a single logistic regression probability
30 score that can be used to classify samples into CAD positive or negative groups. The classifier
31 score is a function of the identified markers that estimates the probability that a subject suffers
32 from CAD (see Materials and Methods for more details). Figure A4-a in the Supplementary
33 Material illustrates the aggregate protein probability score for all samples in the discovery cohort.
34 Figure 3-a shows the receiver operating characteristic curves (ROC) from all cross-validation
35 runs (in light grey) and its average curve (solid blue line) to illustrate the performance of our
36 methodology to classify new samples. Figure A5 in the Supplementary Material illustrates the
37 distribution of the AUCs across the 100 runs using boxplots. Based on these initial results, the
38 identified candidate markers together can predict CAD with a cross-validated sensitivity of 77%,
39 a specificity of 60% and an overall AUC of 0.75 (see Figure 3-a). The estimated sensitivity and
40 specificity in the independent test set, mainly composed of subjects with CAD that died during
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 the followed-up period, was 74% and 55%, respectively.
4
5

6 *Combining the LC/ESI-MRM-MS Results with Clinical Data*

7
8

9 Using the same methodology we built a logistic regression model based on traditional clinical
10 markers of CAD (see Table 1). Figure A4-b in the Supplementary Material illustrates the
11 aggregate protein probability score for all samples in the discovery cohort. The resulting
12 classifier has a performance similar to the proteomics-MRM classifier, with a cross-validated
13 AUC of 0.76, a sensitivity of 80% and a specificity of 57% (Figure 3-b and Figure A5).
14 Similarly, this clinical classifier has a sensitivity of 77% and a specificity of 36% in the test set.
15
16
17
18
19
20
21
22
23

24 Using the proteomics-MRM and the clinical classifiers, we built an ensemble classifier
25 taking the average of their probabilities (see Figure A4-c). Interestingly, the resulting combined
26 classifier has an improved cross-validated AUC of 0.84, a sensitivity of 86% and a specificity of
27 64% (see Figure 3-c and Figure A5). The sensitivity of the ensemble classifier also increased in
28 the test set to 86% compared to that of the clinical and the proteomics-MRM classifiers, and the
29 specificity equals that of the clinical classifier (36%).
30
31
32
33
34
35
36
37
38

39 **Discussion**

40
41

42 Conservative predictions of mortality from CAD are projected to reach 23.4 million people in
43 North America by 2030 [1]. The importance of biomarkers in both diagnosing and determining
44 the prognosis for CAD is well established, with several newer biomarkers being recently
45 identified and applied clinically. This emergence of the diverse range of biomarkers for CAD
46 provides insight into the widely varied pathophysiology of this disease. However, with the
47 increased numbers of biomarkers, there is a critical need for a rapid standardized multiplex
48 measurement platform for diagnosis of CAD and patient-personalized medical treatment
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 interventions. We developed an LC/ESI-MRM-MS assay to measure 100 putative proteins and
4 identified a list of 16 potential proteins associated with CAD (Table 2) that together can stratify
5 subjects with potential risk of CAD (cross-validated AUC of 75%). Combining the proteomics-
6 based MRM classifier with the traditional clinical information about CAD risk factors results in
7 an enhanced classifier to determine the occurrence of CAD among patients at-risk in a non-
8 invasive and timely manner (cross-validated AUC of 84%).
9
10
11
12
13
14
15
16
17

18 Samples from this study were analyzed after a long-term period of storage. The stability
19 of some proteins in these samples was analyzed performing multiple freeze thaw cycles and
20 comparing the resulting protein levels with those from freshly collected plasma [24]. Other
21 studies also have demonstrated the long-term stability of some of the proteins in our study,
22 including CRP [25], SHBG [26] [27], and PON1 [28]. To complement this analysis, we
23 compared clinical measurements of CRP from fresh plasma of 188 patients in this study [14-16]
24 with our MRM-MS measurement from long-term frozen plasma. The estimated robust
25 correlation between the clinical and MRM-MS measures was, on average, 0.96 (see Figure A6 in
26 the Supplementary Material). Considering the intrinsic differences between these technologies,
27 we interpret these results as a favorable indication of sample stability for this inflammatory
28 protein marker.
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43

44 Coronary atherosclerotic plaque formation and disruption with consequent platelet
45 aggregation and thrombosis are the two most important mechanisms that lead to the acute
46 ischemic conditions of CAD [29]. Inflammation is essential to the initiation, development, and
47 progression of atherosclerosis, and therefore it should be no surprise that a number of CAD
48 biomarkers are involved in this inflammatory process. The association of increased serum levels
49 of acute-phase proteins with the progression of atherosclerosis and occurrence of CAD has been
50
51
52
53
54
55
56
57
58
59
60

1
2
3 well documented [30-32]. In particular, C reactive protein (CRP) is an acute phase protein of
4
5 hepatic origin that has been previously associated with increased risk of CAD [31, 32]. CRP
6
7 increases in response to factors released by macrophages, adipocytes and T cells, and activates
8
9 the complement cascade thereby promoting inflammation [33]. Complement C1q subcomponent
10
11 subunit A (C1QA) initiates the complement cascade by binding to and activating C1r and C1s
12
13 proenzymes into complement C1, the first component of the complement system. Recent studies
14
15 have reported that serum C1QA can interact with adiponectin and increased C1q/adiponectin
16
17 ratio correlates with atherosclerosis severity and CAD pathogenesis [34-36]. Complement C3
18
19 plays a central role in the complement cascade and can lead to the initiation of both the classical
20
21 and alternate complement pathways. This multiple complement pathway activation leads to
22
23 increased signal amplification and activation of the membrane attack complex (MAC), which is
24
25 involved in cell lysis and apoptosis. In the case of complement activation on vascular smooth
26
27 muscle cells, this can lead also to cell contraction, increased vascular permeability, and leukocyte
28
29 chemoattraction. Elevated complement C3 levels are associated with the atherosclerotic process
30
31 and arterial calcification, especially in women [37, 38]. Both complement components C1QA
32
33 and C3 were elevated in CAD+ subjects and significantly associated with CAD in our study
34
35 (Table 2 and Figure 1, un-adjusted p-value equal to 0.03 and 0.04, respectively).
36
37
38
39
40
41
42

43 Apolipoprotein L1 (ApoL1) is a minor apolipoprotein component of high-density
44
45 lipoprotein (HDL) and is found in many tissues including the vascular epithelium. The exact role
46
47 of ApoL1 is still being elucidated, but it is thought that ApoL1 may play a role in the
48
49 inflammatory response, as proinflammatory cytokines increase its expression [39]. Similarly,
50
51 apolipoprotein CII (ApoCII), another apolipoprotein, but a component of very low density
52
53 lipoproteins (VLDL) and chylomicrons, is associated with a higher risk of atherosclerosis,
54
55
56
57
58
59
60

1
2
3 although the exact mechanism is still being researched [40]. Therefore, both ApoL1 and ApoCII
4 interact with lipoproteins, such as HDL, LDL, and VLDL, and are involved in the CAD
5 pathogenic process through mechanisms that are still being researched [41]. In our study, CRP
6 and both ApoL1 and ApoCII were significantly elevated in the case group (CAD+) relative to the
7 control group (CAD-) in our study (Table 2 and Figure 1, un-adjusted p-values of 0.01, 0.03, and
8 0.03, respectively) indicating the association of these proteins with CAD. CD5 antigen-like
9 (CD5L) is a secreted protein mainly expressed by macrophages in inflamed tissues that acts a key
10 regulator of lipids and is responsible for progression of the atherosclerotic process in CAD [42].
11 CD5L is a potent inhibitor of apoptosis in macrophages and promotes macrophage survival from
12 the apoptotic effects of oxidized lipids [42]. CD5L was, on average, more abundant in CAD+
13 subjects relative to CAD- subjects in our study, reflecting its association with inflammatory
14 process in CAD (Table 2 and Figure 1, un-adjusted p-value of 0.02).

15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32 Paraoxonases are a family of mammalian enzymes that have important anti-inflammatory,
33 anti-oxidative, anti-atherogenic, anti-diabetic, anti-microbial, and organophosphate-hydrolyzing
34 properties [43, 44]. Paraoxonase 1 (PON1) is primarily synthesized in the liver and is transported
35 into the bloodstream where it binds with high-density lipoprotein (HDL). PON1 plays an
36 important role in preventing the oxidation of low-density lipoproteins (LDL), which is directly
37 implicated in the development of atherosclerosis. PON1 polymorphisms and changes in
38 biological activity can impact lipid metabolism, leading to the aggravation of atherosclerosis and
39 increase the risk of a coronary event [45-47]. Furthermore, PON1 serum concentration is
40 involved in a feedback loop and greatly influenced by inflammation and serum oxidized (ox)-
41 LDL levels. In our study, serum PON1 was significantly associated with CAD (Table 2 and
42 Figure 1, un-adjusted p-value of 0.03).

1
2
3 Coagulation factor X (F10) is the first member of the final common pathway for
4 thrombosis and, when activated, converts large amounts of prothrombin (inactive form) into
5 thrombin (active) on the surface of activated platelets. This process, known as the “thrombin
6 burst”, is responsible for fibrin polymerization that leads to fibrin clot formation and thrombosis.
7
8 Therefore, F10 is a critical positive mediator of thrombosis, and is the target of commonly used
9 anti-coagulants in clinical practice, such as warfarin and heparin. Activated coagulation factor V
10 (F5) can bind to activated platelets and act as a cofactor for activated F10, facilitating “thrombin
11 burst”, fibrin clot formation, and thrombosis. Furthermore, fibronectin, an extracellular matrix
12 glycoprotein, can bind to the fibrin clot and plays a key role in cell adhesion, growth, migration,
13 and differentiation, thereby enhancing the role of macrophages and fibroblasts in fibrin clot
14 stabilization and thrombosis [48]. In our analysis, F10 and F5, were significantly elevated in
15 subjects with CAD (Table 2 and Figure 1, un-adjusted p-values of 0.01 and 0.04, respectively).
16
17 In contrast, fibronectin levels were significantly decreased in case subjects, as compared to
18 controls (Table 2 and Figure 1, un-adjusted p-value < 0.001). Although this process may be due
19 to enhanced fibrinosis and the incorporation of soluble plasma fibrinogen into the insoluble blood
20 clot during CAD, we have not detected significant changes of fibrinogen in this study. von
21 Willebrand factor (vWF), a protein secreted by endothelial cells, is important in the formation
22 and maintenance of a fibrin clot. Specifically, during thrombosis, it promotes adhesion of
23 platelets to the fibrin clot by forming a molecular bridge between sub-endothelial collagen matrix
24 and platelet-surface receptor complexes. Coagulation factor XIII (FXIIIa) can also covalently
25 cross-link vWF to polymerizing fibrin and fibrin-bound VWF further recruits and activates
26 platelets to support fibrin clot formation [49]. Though an exact mechanism has not been
27 elucidated, both vWF and fibronectin are known to support the formation of the fibrin clot and
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 have been shown to be co-localized in the blood vessel endothelium during cardiovascular
4 disease [50, 51]. High vWF levels are predictive of adverse cardiovascular outcomes [52] and, in
5 agreement with these previous findings, vWF was elevated in our case subjects when compared
6 to control subjects (Table 2 and Figure 1, un-adjusted p-value < 0.001).
7
8
9
10
11
12

13 Fibrinolysis is a process that breaks down and prevents fibrin clots, the end product of
14 coagulation, from growing and becoming problematic [53]. Plasmin, the key effector for
15 fibrinolysis, is inactivated by alpha 2-antiplasmin (A2AP) [54, 55]. Furthermore, A2AP has a key
16 role in regulating cytokine signaling and phagocytosis activity during inflammation [56], both of
17 which are implicated in the pathogenic process of CAD. Thus, the significantly elevated levels of
18 A2AP in CAD+ subjects (Table 2 and Figure 1, un-adjusted p-value of 0.01) are not necessarily
19 surprising.
20
21
22
23
24
25
26
27
28
29

30 Protein S is a vitamin K-dependent plasma glycoprotein synthesized by the endothelium
31 and involved in the blood coagulation cascade. Protein S is partially homologous to other
32 vitamin-K dependent proteins, including FX, and to plasma steroid hormone-binding proteins
33 such as sex hormone-binding globulin (SHBG). Protein S is best characterized in its role in the
34 anti-coagulation pathway, where it functions as a cofactor of protein C in the inactivation of
35 coagulation factors, such as activated FV [57-59]. Specifically, activated FV is degraded by
36 protein C, one of the principal physiological inhibitors of coagulation, which is activated by
37 thrombin. In other words, thrombin limits its own activation in a negative feedback loop via the
38 concentration levels of activated protein C. We also found a significantly higher level of protein
39 S in the CAD+ group, relative to CAD-, in our cohort (Table 2 and Figure, un-adjusted p-value <
40 0.0001, respectively). The significantly elevated levels of pro-inflammatory markers are a strong
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 indicator of the role inflammation plays in atherosclerotic plaque formation and thrombosis in
4
5 CAD.
6
7

8
9 Sex hormone-binding globulin (SHBG) is a glycoprotein that is produced and secreted by
10 the liver into the bloodstream where it binds sex steroids, including testosterone and estradiol.
11 Testosterone and estradiol circulate in the bloodstream, bound mostly to SHBG, where these
12 bound hormones remain functionally inactive. Thus, bioavailability of sex hormones is
13 influenced by the level of SHBG. Furthermore, the proinflammatory mechanisms that are
14 implicated in CAD pathogenesis through increased cytokine signaling and phagocytosis activity
15 are known to also regulate SHBG activity [60]. Therefore, the cortisol/testosterone and
16 cortisol/SHBG ratios are hypothesized to be strong predictors of atherosclerosis [61], and our
17 results support the hypothesis that low SHBG is a strong predictor for CAD, with significant
18 lower levels in the CAD+ group (Table 2 and Figure 1, un-adjusted p-value of 0.04).
19
20
21
22
23
24
25
26
27
28
29
30
31

32 Iron status is a modifiable property that has been implicated in a number of
33 cardiovascular diseases, and higher iron status reduces CAD risk [62]. Specifically, both low
34 hemoglobin and iron depletion are independently associated with CAD [63]. Transferrins are
35 iron-binding transport proteins that are responsible for the transport of iron from absorption and
36 heme degradation sites to storage and utilization sites. Significantly increased levels of
37 transferrin can lead to low hemoglobin and iron depletion and may exacerbate CAD
38 development. In our study, transferrin was shown to also be associated with CAD (Table 2 and
39 Figure 1, un-adjusted p-value of 0.01). PON1 has important protective effects in cardiovascular
40 diseases, including regulating cytokine signaling and phagocytosis [64, 65]. PON1 is also
41 associated with regulating iron metabolism and maintaining proper iron homeostasis [66] and
42 iron dysregulation has been implicated as a risk factor for CAD [67, 68]. As indicated previously,
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 in our study, serum PON1 was significantly associated with CAD (Table 2 and Figure 1, un-
4 adjusted p-value of 0.03).
5
6

7
8 Most of our 16 discriminatively protein biomarkers are primarily involved in the two
9 most important early mechanisms for CAD pathogenesis: (i) atherosclerotic plaque formation
10 and (ii) thrombosis. Furthermore, we have uncovered other associated proteins that are indirectly
11 involved in the pathogenic processes, such as in the regulation of cytokines and phagocytic cells,
12 or interactions with lipoproteins, or in maintaining iron homeostasis. Thus, the identified panel
13 can be used for a better stratification of at-risk CAD patients with evidence of a more thrombotic
14 and/or pro-inflammatory extracellular environment. The translation of these markers to the clinic
15 would still require an extensive external validation and a migration of the markers to a clinical
16 assay [8].
17
18
19
20
21
22
23
24
25
26
27
28
29

30 Overall, the results of this study illustrate the potential of using an accurate and
31 reproducible LC/ESI-MRM-MS assay to discover promising proteins that can be combined with
32 usual clinical risk factor to enhance risk evaluation of CAD.
33
34
35
36
37
38
39
40

41 **Acknowledgments**

42
43

44 We thank the University of Victoria Genome BC Proteomics Centre for their expertise in
45 conducting the LC/ESI-MRM-MS experiments, and in particular Carol Parker for their
46 comments on this manuscript. This work was funded by the NCE Centre of Excellence for
47 Commercialization and Research (McManus et al). The University of Victoria - Genome BC
48 Proteomics Centre is funded by the Genome Canada and Genome British Columbia funds to
49 (project codes 204PRO for operations and 214PRO for technology development). GAF was
50
51
52
53
54
55
56
57
58
59
60

1
2
3 funded by CIHR Project Grant PJT 148883. GCF is funded by an NSERC Discovery Grant. CHB
4
5 is grateful for support from the Leading Edge Endowment Fund (LEEF). CHB is also grateful
6
7 for support from the Segal McGill Chair in Molecular Oncology at McGill University (Montreal,
8
9 Quebec, Canada), and for support from the Warren Y. Soper Charitable Trust and the Alvin
10
11 Segal Family Foundation to the Jewish General Hospital (Montreal, Quebec, Canada).
12
13

14
15
16 CHB is the CSO of MRM Proteomics, Inc. All other authors have no financial/commercial
17
18 conflicts of interests.
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

References

- [1] Mozaffarian, D., Benjamin, E. J., Go, A. S., Arnett, D. K., *et al.*, Heart disease and stroke statistics--2015 update: a report from the American Heart Association. *Circulation* 2015, *131*, e29-322.
- [2] Yusuf, S., Hawken, S., Ounpuu, S., Dans, T., *et al.*, Effect of potentially modifiable risk factors associated with myocardial infarction in 52 countries (the INTERHEART study): case-control study. *Lancet* 2004, *364*, 937-952.
- [3] O'Donnell, M. J., Xavier, D., Liu, L., Zhang, H., *et al.*, Risk factors for ischaemic and intracerebral haemorrhagic stroke in 22 countries (the INTERSTROKE study): a case-control study. *Lancet* 2010, *376*, 112-123.
- [4] Libby, P., Theroux, P., Pathophysiology of coronary artery disease. *Circulation* 2005, *111*, 3481-3488.
- [5] Kannel, W. B., Feinleib, M., Natural history of angina pectoris in the Framingham study. Prognosis and survival. *Am J Cardiol* 1972, *29*, 154-163.
- [6] Kumar, A., Cannon, C. P., Acute coronary syndromes: diagnosis and management, part I. *Mayo Clin Proc* 2009, *84*, 917-938.
- [7] Anderson, L., Candidate-based proteomics in the search for biomarkers of cardiovascular disease. *J Physiol* 2005, *563*, 23-60.
- [8] Cohen Freue, G. V., Meredith, A., Smith, D., Bergman, A., *et al.*, Computational biomarker pipeline from discovery to clinical implementation: plasma proteomic biomarkers for cardiac transplantation. *PLoS Comput Biol* 2013, *9*, e1002963.
- [9] Bermudez-Crespo, J., Lopez, J. L., A better understanding of molecular mechanisms underlying human disease. *Proteomics Clin Appl* 2007, *1*, 983-1003.

- 1
2
3 [10] Kondrat, R. W., McClusky, G. A., Cooks, R. G., Multiple reaction monitoring in mass
4 spectrometry/mass spectrometry for direct analysis of complex mixtures. *Analytical Chemistry*
5 1978, *50*, 2017-2021.
6
7
8
9
10 [11] Kiyonami, R., Schoen, A., Prakash, A., Peterman, S., *et al.*, Increased selectivity, analytical
11 precision, and throughput in targeted proteomics. *Mol Cell Proteomics* 2011, *10*, M110 002931.
12
13 [12] Gergov, M., Ojanpera, I., Vuori, E., Simultaneous screening for 238 drugs in blood by liquid
14 chromatography-ion spray tandem mass spectrometry with multiple-reaction monitoring. *J*
15 *Chromatogr B Analyt Technol Biomed Life Sci* 2003, *795*, 41-53.
16
17
18
19 [13] Günther, O. P., Chen, V., Freue, G. C., Balshaw, R. F., *et al.*, A computational pipeline for
20 the development of multi-marker bio-signature panels and ensemble classifiers. *BMC Bioinform.*
21 2012, *13*, 326.
22
23 [14] Hill, J., Heslop, C., Man, S. F., Frohlich, J., *et al.*, Circulating surfactant protein-D and the
24 risk of cardiovascular morbidity and mortality. *Eur Heart J* 2011, *32*, 1918-1925.
25
26 [15] Heslop, C. L., Frohlich, J. J., Hill, J. S., Myeloperoxidase and C-reactive protein have
27 combined utility for long-term prediction of cardiovascular mortality after coronary angiography.
28 *J Am Coll Cardiol* 2010, *55*, 1102-1109.
29
30 [16] Lee, K. W., Hill, J. S., Walley, K. R., Frohlich, J. J., Relative value of multiple plasma
31 biomarkers as risk factors for coronary artery disease and death in an angiography cohort. *CMAJ*
32 2006, *174*, 461-466.
33
34 [17] Kuzyk, M. A., Smith, D., Yang, J., Cross, T. J., *et al.*, Multiple reaction monitoring-based,
35 multiplexed, absolute quantitation of 45 proteins in human plasma. *Mol Cell Proteomics* 2009, *8*,
36 1860-1877.
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

- 1
2
3 [18] Spitzer, A. R., Chace, D., Mass spectrometry in neonatal medicine and clinical diagnosis--
4 the [corrected] potential use of mass spectrometry in neonatal brain [corrected] monitoring. *Clin*
5 *Perinatol* 2006, *33*, 729-744, viii.
6
7
8
9
10 [19] Hollander, Z., Lazarova, M., Lam, K. K., Ignaszewski, A., *et al.*, Proteomic biomarkers of
11 recovered heart function. *Eur J Heart Fail* 2014, *16*, 551-559.
12
13
14 [20] Shmueli, G., To explain or to predict? *Statistical science* 2010, *25*, 289-310.
15
16
17 [21] Sing, T., Sander, O., Beerenwinkel, N., Lengauer, T., ROCR: visualizing classifier
18 performance in R. *Bioinformatics* 2005, *21*, 3940-3941.
19
20
21 [22] LeDell, E., Petersen, M., van der Laan, M., Computationally efficient confidence intervals
22 for cross-validated area under the ROC curve estimates. *Electronic Journal of Statistics* 2015, *9*,
23 1583-1607.
24
25
26
27
28 [23] Domanski, D., Percy, A. J., Yang, J., Chambers, A. G., *et al.*, MRM-based multiplexed
29 quantitation of 67 putative cardiovascular disease biomarkers in human plasma. *Proteomics*
30 2012, *12*, 1222-1243.
31
32
33
34
35 [24] Heslop, C. L., Tebbutt, S. J., Podder, M., Ruan, J., Hill, J. S., Combined polymorphisms in
36 oxidative stress genes predict coronary artery disease and oxidative stress in coronary
37 angiography patients. *Ann Hum Genet* 2012, *76*, 435-447.
38
39
40
41
42 [25] Doumatey, A. P., Zhou, J., Adeyemo, A., Rotimi, C., High sensitivity C-reactive protein
43 (Hs-CRP) remains highly stable in long-term archived human serum. *Clin Biochem* 2014, *47*,
44 315-318.
45
46
47
48
49 [26] Stroud, L. R., Solomon, C., Shenassa, E., Papandonatos, G., *et al.*, Long-term stability of
50 maternal prenatal steroid hormones from the National Collaborative Perinatal Project: still valid
51 after all these years. *Psychoneuroendocrinology* 2007, *32*, 140-150.
52
53
54
55
56
57
58
59
60

- 1
2
3 [27] Gislefoss, R. E., Grimsrud, T. K., Morkrid, L., Stability of selected serum proteins after
4 long-term storage in the Janus Serum Bank. *Clin Chem Lab Med* 2009, *47*, 596-603.
5
6
7 [28] Beekhof, P. K., Gorshunska, M., Jansen, E. H., Long term stability of paraoxonase-1 and
8 high-density lipoprotein in human serum. *Lipids Health Dis* 2012, *11*, 53.
9
10
11 [29] Ross, R., Atherosclerosis--an inflammatory disease. *N Engl J Med* 1999, *340*, 115-126.
12
13 [30] Berk, B. C., Weintraub, W. S., Alexander, R. W., Elevation of C-reactive protein in "active"
14 coronary artery disease. *The American Journal of Cardiology* 1990, *65*, 168-172.
15
16
17 [31] Lindahl, B., Toss, H., Siegbahn, A., Venge, P., Wallentin, L., Markers of myocardial
18 damage and inflammation in relation to long-term mortality in unstable coronary artery disease.
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
- FRISC Study Group. Fragmin during Instability in Coronary Artery Disease. *N Engl J Med* 2000,
343, 1139-1147.
- [32] Rossi, E., Biasucci, L. M., Citterio, F., Pelliccioni, S., *et al.*, Risk of myocardial infarction
and angina in patients with severe peripheral vascular disease: predictive role of C-reactive
protein. *Circulation* 2002, *105*, 800-803.
- [33] Kubkova, L., Spinar, J., Pavkova Goldbergova, M., Jarkovsky, J., Parenica, J.,
[Inflammatory response and C -reactive protein value in patient with acute coronary syndrome].
Vnitr Lek 2013, *59*, 981-988.
- [34] Hong, E. S., Lim, C., Choi, H. Y., Ku, E. J., *et al.*, The amount of C1q-adiponectin complex
is higher in the serum and the complex localizes to perivascular areas of fat tissues and the
intimal-medial layer of blood vessels of coronary artery disease patients. *Cardiovasc Diabetol*
2015, *14*, 50.

- 1
2
3 [35] Hirata, A., Kishida, K., Nakatsuji, H., Kobayashi, H., *et al.*, High serum C1q-
4 adiponectin/total adiponectin ratio correlates with coronary artery disease in Japanese type 2
5 diabetics. *Metabolism* 2013, *62*, 578-585.
6
7
8
9
10 [36] Nakatsuji, H., Kobayashi, H., Kishida, K., Nakagawa, T., *et al.*, Binding of adiponectin and
11 C1q in human serum, and clinical significance of the measurement of C1q-adiponectin / total
12 adiponectin ratio. *Metabolism* 2013, *62*, 109-120.
13
14
15
16
17 [37] Nagaraj, N., Matthews, K. A., Shields, K. J., Barinas-Mitchell, E., *et al.*, Complement
18 proteins and arterial calcification in middle aged women: Cross-sectional effect of cardiovascular
19 fat. The SWAN Cardiovascular Fat Ancillary Study. *Atherosclerosis* 2015, *243*, 533-539.
20
21
22
23
24 [38] Snyder, M. L., Shields, K. J., Korytkowski, M. T., Sutton-Tyrrell, K., Talbott, E. O.,
25 Complement protein C3 and coronary artery calcium in middle-aged women with polycystic
26 ovary syndrome and controls. *Gynecol Endocrinol* 2014, *30*, 511-515.
27
28
29
30
31 [39] Wan, G., Zhaorigetu, S., Liu, Z., Kaini, R., *et al.*, Apolipoprotein L1, a novel Bcl-2
32 homology domain 3-only lipid-binding protein, induces autophagic cell death. *J Biol Chem* 2008,
33 *283*, 21540-21549.
34
35
36
37
38 [40] Li, Y., He, P. P., Zhang, D. W., Zheng, X. L., *et al.*, Lipoprotein lipase: from gene to
39 atherosclerosis. *Atherosclerosis* 2014, *237*, 597-608.
40
41
42
43 [41] Ding, D., Li, X., Qiu, J., Li, R., *et al.*, Serum lipids, apolipoproteins, and mortality among
44 coronary artery disease patients. *Biomed Res Int* 2014, *2014*, 709756.
45
46
47 [42] Amezaga, N., Sanjurjo, L., Julve, J., Aran, G., *et al.*, Human scavenger protein AIM
48 increases foam cell formation and CD36-mediated oxLDL uptake. *J Leukoc Biol* 2014, *95*, 509-
49 520.
50
51
52
53
54
55
56
57
58
59
60

- 1
2
3 [43] Bergmeier, C., Siekmeier, R., Gross, W., Distribution spectrum of paraoxonase activity in
4 HDL fractions. *Clin Chem* 2004, *50*, 2309-2315.
5
6
7 [44] Aggarwal, G., Prajapati, R., Tripathy, R. K., Bajaj, P., *et al.*, Toward Understanding the
8 Catalytic Mechanism of Human Paraoxonase 1: Site-Specific Mutagenesis at Position 192. *PLoS*
9 *One* 2016, *11*, e0147999.
10
11
12 [45] Martinez-Quintana, E., Rodriguez-Gonzalez, F., Medina-Gil, J. M., Garay-Sanchez, P.,
13 Tugores, A., Paraoxonase 1 (Q192R) gene polymorphism, coronary heart disease and the risk of
14 a new acute coronary event. *Clin Investig Arterioscler* 2017, *29*, 1-6.
15
16
17 [46] Fridman, O., Gariglio, L., Riviere, S., Porcile, R., *et al.*, Paraoxonase 1 gene polymorphisms
18 and enzyme activities in coronary artery disease and its relationship to serum lipids and
19 glycemia. *Arch Cardiol Mex* 2016, *86*, 350-357.
20
21
22 [47] Hernandez-Diaz, Y., Tovilla-Zarate, C. A., Juarez-Rojop, I. E., Gonzalez-Castro, T. B., *et*
23 *al.*, Effects of paraoxonase 1 gene polymorphisms on heart diseases: Systematic review and
24 meta-analysis of 64 case-control studies. *Medicine (Baltimore)* 2016, *95*, e5298.
25
26
27 [48] Williams, C. M., Engler, A. J., Slone, R. D., Galante, L. L., Schwarzbauer, J. E., Fibronectin
28 expression modulates mammary epithelial cell proliferation during acinar differentiation. *Cancer*
29 *Res* 2008, *68*, 3185-3192.
30
31
32 [49] Miszta, A., Pelkmans, L., Lindhout, T., Krishnamoorthy, G., *et al.*, Thrombin-dependent
33 Incorporation of von Willebrand Factor into a Fibrin Network. *J Biol Chem* 2014, *289*, 35979-
34 35986.
35
36
37 [50] Kloczko, J., Wojtukiewicz, M. Z., Bielawiec, M., Galar, M., Tarasow, E., von Willebrand
38 factor antigen and fibronectin in essential hypertension. *Thromb Res* 1995, *79*, 331-336.
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

- 1
2
3 [51] Solerte, S. B., Piovela, F., Viola, C., Schianca, G. C., *et al.*, Plasma fibronectin, von
4 Willebrand factor antigen, and blood rheology. Association with diabetic microvascular disease.
5
6 *Acta Diabetol Lat* 1985, 22, 239-246.
7
8
9
10 [52] Sonneveld, M. A., Cheng, J. M., Oemrawsingh, R. M., de Maat, M. P., *et al.*, Von
11 Willebrand factor in relation to coronary plaque characteristics and cardiovascular outcome.
12 Results of the ATHEROREMO-IVUS study. *Thromb Haemost* 2015, 113, 577-584.
13
14
15
16 [53] Cesarman-Maus, G., Hajjar, K. A., Molecular mechanisms of fibrinolysis. *Br J Haematol*
17 2005, 129, 307-321.
18
19
20
21 [54] Shieh, B. H., Travis, J., The reactive site of human alpha 2-antiplasmin. *J Biol Chem* 1987,
22 262, 6055-6059.
23
24
25
26 [55] Wiman, B., Collen, D., On the mechanism of the reaction between human alpha 2-
27 antiplasmin and plasmin. *J Biol Chem* 1979, 254, 9291-9297.
28
29
30
31 [56] Kager, L. M., Weehuizen, T. A., Wiersinga, W. J., Roelofs, J. J., *et al.*, Endogenous alpha2-
32 antiplasmin is protective during severe gram-negative sepsis (melioidosis). *Am J Respir Crit*
33 *Care Med* 2013, 188, 967-975.
34
35
36
37 [57] Heeb, M. J., Kojima, Y., Rosing, J., Tans, G., Griffin, J. H., C-terminal residues 621-635 of
38 protein S are essential for binding to factor Va. *J Biol Chem* 1999, 274, 36187-36192.
39
40
41
42 [58] Heeb, M. J., Mesters, R. M., Tans, G., Rosing, J., Griffin, J. H., Binding of protein S to
43 factor Va associated with inhibition of prothrombinase that is independent of activated protein C.
44
45
46
47 *J Biol Chem* 1993, 268, 2872-2877.
48
49
50 [59] Castoldi, E., Hackeng, T. M., Regulation of coagulation by protein S. *Curr Opin Hematol*
51 2008, 15, 529-536.
52
53
54
55
56
57
58
59
60

- 1
2
3 [60] Simo, R., Saez-Lopez, C., Barbosa-Desongles, A., Hernandez, C., Selva, D. M., Novel
4 insights in SHBG regulation and clinical implications. *Trends Endocrinol Metab* 2015, 26, 376-
5 383.
6
7
8
9
10 [61] Lee, J. M., Colangelo, L. A., Schwartz, J. E., Yano, Y., *et al.*, Associations of
11 cortisol/testosterone and cortisol/sex hormone-binding globulin ratios with atherosclerosis in
12 middle-age women. *Atherosclerosis* 2016, 248, 203-209.
13
14
15
16 [62] Gill, D., Del Greco, M. F., Walker, A. P., Srai, S. K. S., *et al.*, The Effect of Iron Status on
17 Risk of Coronary Artery Disease: A Mendelian Randomization Study. *Arterioscler Thromb Vasc*
18 *Biol* 2017.
19
20
21
22
23 [63] Grammer, T. B., Kleber, M. E., Silbernagel, G., Pilz, S., *et al.*, Hemoglobin, iron
24 metabolism and angiographic coronary artery disease (The Ludwigshafen Risk and
25 Cardiovascular Health Study). *Atherosclerosis* 2014, 236, 292-300.
26
27
28
29 [64] Martinelli, N., Consoli, L., Girelli, D., Grison, E., *et al.*, Paraoxonases: ancient substrate
30 hunters and their evolving role in ischemic heart disease. *Adv Clin Chem* 2013, 59, 65-100.
31
32
33 [65] Litvinov, D., Mahini, H., Garelnabi, M., Antioxidant and anti-inflammatory role of
34 paraoxonase 1: implication in arteriosclerosis diseases. *N Am J Med Sci* 2012, 4, 523-532.
35
36
37 [66] Okuturlar, Y., Gedikbasi, A., Akalin, N., Gunaldi, M., *et al.*, Serum paraoxonase 1 activity
38 in patients with iron deficiency anemia. *Arch Med Sci* 2016, 12, 697-703.
39
40
41 [67] Spasojevic-Kalimanovska, V., Bogavac-Stanojevic, N., Kalimanovska-Ostic, D., Memon,
42 L., *et al.*, Factor analysis of risk variables associated with iron status in patients with coronary
43 artery disease. *Clin Biochem* 2014, 47, 564-569.
44
45
46 [68] Bagheri, B., Shokrzadeh, M., Mokhberi, V., Azizi, S., *et al.*, Association between Serum
47 Iron and the Severity of Coronary Artery Disease. *Int Cardiovasc Res J* 2013, 7, 95-98.
48
49
50
51
52
53
54
55
56
57
58
59
60

Figure Legends

Figure 1. Boxplot plots of candidate biomarker proteins in the derived panel. Points represent the log-transformed (base 2) standardized protein levels.

Figure 2. Pairwise robust correlations between all peptides from the 16 identified candidate biomarker proteins. Peptide names are shown as row names (left axis). Protein acronyms for each proteotypic peptide are shown as the corresponding column name (bottom axis).

Figure 3. Receiver operating characteristic curves of the (a) proteomics-MRM, (b) the clinical, and (c) the ensemble classifiers estimated by 100 runs 5-fold cross-validations (grey lines). The blue curve represents the average of the 100 ROC curves estimated within each cross-validation run. The horizontal line represents the sensitivity of the classifier using a probability cut-off of 0.5.

Table 1. Description of the discovery cohort.

	CAD- (n=45)	CAD+ (n=70)	un- adjusted p-value
Previous myocardial infarction, %	11.1	42.9	<0.001
Diabetes mellitus, %	4.4	22.9	0.02
Previous angioplasty, %	2.2	15.7	0.05
Previous bypass, %	0.0	11.4	0.05
Family history of CAD, %	46.6	60.0	0.23
Hypertension, %	40.0	41.4	0.99
Smoking, %	55.6	73.9	0.07
Apolipoprotein B, g/L	0.9 ± 0.2	1.0 ± 0.2	0.02
Body mass index, kg/m²	26.6 ± 3.2	28.1 ± 3.8	0.03
TC:HDL cholesterol ratio	5.1 ± 1.6	5.8 ± 1.7	0.05
Waist circumference, cm	87.2 ± 29.1	97.7 ± 23.1	0.03
LDL cholesterol, mmol/L	3.3 ± 0.9	3.5 ± 1.0	0.24
HDL cholesterol, mmol/L	1.0 ± 0.2	0.9 ± 0.2	0.16

Table 2. List of panel proteins with significant differential abundance in subjects with (CAD+) and without (CAD-) angiographic evidence of CAD. Average, standard deviations (SD) and fold changes are computed based on the raw relative responses of each selected peptide.

Proteins	Gene Symbol	Accession Number	Average (SD) for CAD+	Average (SD) for CAD-	Fold Change	p-value
Fibronectin	FN1	P02751	0.120 (0.079)	0.190 (0.097)	1.548	<0.001
Protein S	PROS1	P07225	0.052 (0.012)	0.043 (0.010)	1.202	<0.001
von Willebrand Factor	vWF	P04275	0.002 (0.003)	0.005 (0.005)	2.085	0.001
Inter alpha trypsin inhibitor heavy chain H1	ITIH1	P19827	0.250 (0.047)	0.220 (0.040)	1.129	0.001
Coagulation Factor X	F10	P00742	0.030 (0.006)	0.027 (0.006)	1.120	0.009
Alpha 2 antiplasmin	A2AP	P08697	0.040 (0.011)	0.035 (0.008)	1.157	0.010
Transferrin	TF	P02787	0.260 (0.053)	0.240 (0.043)	1.098	0.013
C reactive protein CRP	CRP	P02741	0.008 (0.013)	0.007 (0.016)	1.268	0.014
CD5 antigen-like	CD5L	O43866	0.180 (0.100)	0.140 (0.053)	1.297	0.016
Complement C1 Q subcomponent subunit A	C1QA	P02745	0.003 (0.001)	0.002 (0.001)	1.183	0.025
Apolipoprotein L1	APOL1	O14791	0.043 (0.012)	0.038 (0.013)	1.128	0.026
Serum paraoxonase 1	PON1	P27169	0.130 (0.040)	0.120 (0.031)	1.147	0.031
Apolipoprotein C II	APOC2	P02655	0.088 (0.045)	0.071 (0.046)	1.225	0.035
Complement C3	C3	P01024	0.048 (0.012)	0.044 (0.010)	1.106	0.038

Coagulation Factor V	F5	P12259	0.005 (0.001)	0.004 (0.001)	1.118	0.039
Sex hormone binding globulin	SHBG	P04278	0.009 (0.006)	0.011 (0.007)	1.237	0.042

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

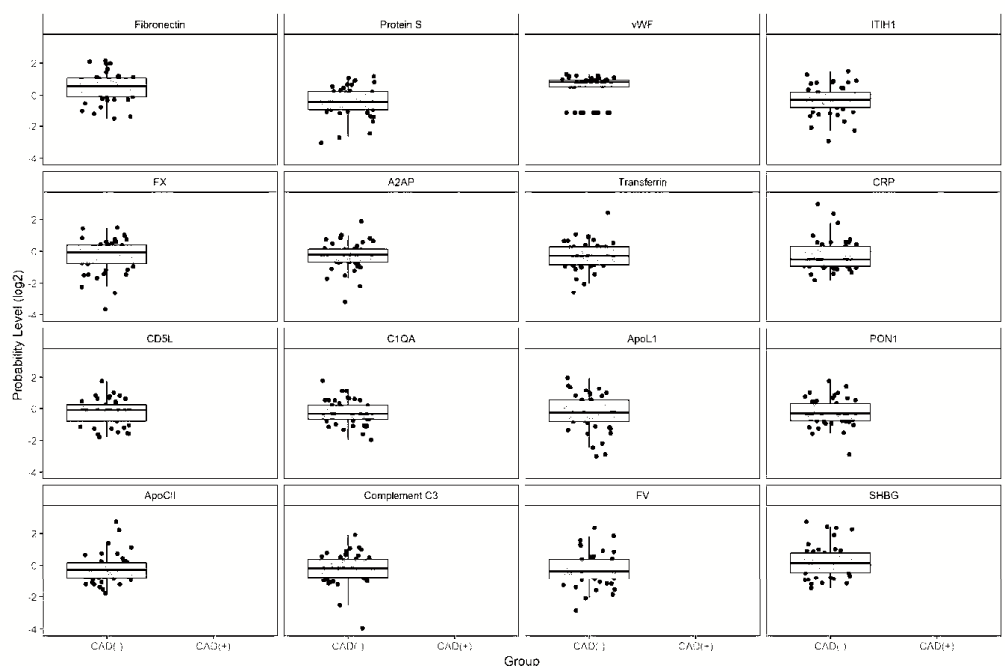


Figure 1. Boxplot plots of candidate biomarker proteins in the derived panel. Points represent the log-transformed (base 2) standardized protein levels.

304x203mm (300 x 300 DPI)

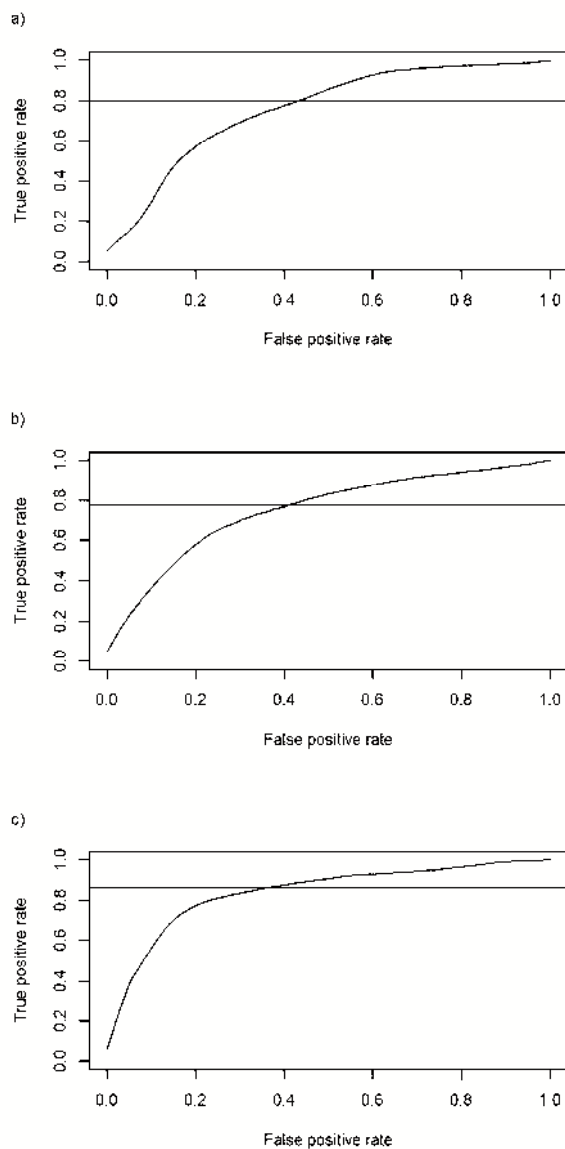


Figure 3. Receiver operating characteristic curves of the (a) proteomics-MRM, (b) the clinical, and (c) the ensemble classifiers estimated by 100 runs 5-fold cross-validations (grey lines). The blue curve represents the average of the 100 ROC curves estimated within each cross-validation run. The horizontal line represents the sensitivity of the classifier using a probability cut-off of 0.5.

101x203mm (300 x 300 DPI)

Reviewer: 1**Comments to the Author**

The central hypothesis of the presented manuscript is that novel protein in plasma can serve as predictors of cardiovascular risk. The authors apply targeted proteomics panel of 100 proteins developed at U Victoria to plasma from two sets – one, a discovery, model building case-control cohort of 45 people with clinical CVD and 70 of people without clinical CVD (defined by angiography). The discovery set is also used to test the model by applying a form of leave-one-out five-fold cross-validation approach. The model comprising of 16 proteins (most upregulated with CVD) is then applied to a separate cohort of subjects (n=87) of which most died of CVD in the follow up to blood collection (n=73). The model is also compared to a model built using common risk factors and the authors show that the proteomics predictor performs about the same as common risk predictor and when combined, the risk prediction is improved. There are several problems with the presented study –

We thank the reviewer for the insightful comments and suggestions, which generated an improved presentation of our work and a more robust analysis of our data. Please find a point-to-point response to your comments.

- 1. The model is developed with a population where cases already have clinical CVD and is not properly validated in a population which includes both healthy and people with CVD.*

Both training and test set are from a cohort of subjects who underwent coronary angiography for suspected coronary artery disease (CAD). We recognize that this study does not involve a group of healthy controls. Thus, our goal is to enhanced risk stratification and diagnosis of CAD among an at-risk population of patients.

To clarify this important point raised by the reviewer, we have reviewed the last sentence of the Introduction (page 6) to:

“The results of this study are a solid first step toward the use of multiplexed LC/ESI-MRM-MS assays in proteomic discovery studies, and for the development of a robust and accurate tool to determine who among an at-risk population has developed clinical CAD.”

We also reviewed the following sentences of the Discussion:

Page 19: “We developed an LC/ESI-MRM-MS assay to measure 100 putative proteins and identified a list of 16 potential proteins associated with CAD (Table 2) that together can stratify subjects with potential risk of CAD (cross-validated AUC of 75%). Combining the proteomics-based MRM classifier with the traditional clinical information about CAD risk factors results in an enhanced classifier to determine the occurrence of CAD among patients at-risk in a non-invasive and timely manner (cross-validated AUC of 84%).”

Page 25: “The results of this study illustrate the potential of using an accurate and reproducible LC/ESI-MRM-MS assay to discover promising proteins that can be used for risk evaluation of CAD.”

1
2
3
4 Similar statements throughout the paper emphasize that the aim of our study is to evaluate
5 patients with suspected CAD (e.g., last 2 sentences of the Statement of clinical relevance section,
6 conclusions of the Abstract, and second and third paragraphs of the Introduction).
7

8
9 As noted by the reviewer, the independent test set is mainly composed of subjects with CAD that
10 died during the followed-up period. Thus, we've also evaluated the proposed classifier using
11 cross-validation, which creates a valid test set from the discovery cohort of patients. In a 5-fold
12 cross-validation, the discovery cohort is divided into 5 sets (folds). Each of these sets is left out
13 as a test set and the remaining 4 sets are combined to form a training set. Importantly, the test set
14 (fold left out) is not used to identify potential markers or build the classifier but only to estimate
15 the performance of the resulting model. The advantage of using a cross-validation is that the
16 population of the test set used resembles that of the training set without overestimating the
17 performance of the classifier. We now highlight and clarify this problem in the second paragraph
18 of page 14 in the Materials and Methods-Statistical Analysis section.
19
20

21 Noteworthy, similar performance results were obtained from both validations: the cross-validated
22 sensitivity and specificity are 77% and 60%, respectively; the estimated sensitivity and
23 specificity in the independent test set are 74% and 55%, respectively.
24
25

26 2. *It is not clear how old the samples are – indication is at least 6 years, but cohort description*
27 *suggests it could be as long as 25 years. Were all samples of the same age? Is there data*
28 *suggesting sample stability?*
29

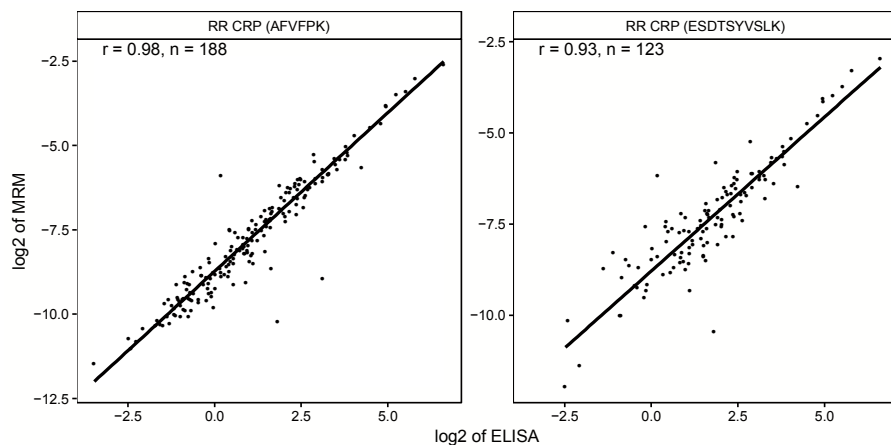
30
31 As indicated in Materials and Methods, the analyzed cohort is comprised of men and women
32 who were referred to two major teaching hospitals in Vancouver for selective coronary
33 angiography between 1992 and 1995. Blood samples were collected just before the angiography
34 was performed.
35

36 Plasma lipid levels, apolipoprotein B and C-reactive protein (CRP) levels were immediately
37 measured at time of sample collection (see references [14-16] of the manuscript). Blood samples
38 were centrifuged immediately and aliquoted into smaller vials and frozen in -70°C until
39 analysis. In their paper from the Annals of Human Genetics (2012, Vol 76, pages 435–447), Dr.
40 Hill's group analyzed the stability of the biomarker proteins identified in their study performing
41 multiple freeze thaw cycles and comparing the resulting protein levels with those from freshly
42 collected plasma. Although the plasma samples analyzed by Dr. Hill in 2012 are the same as
43 those used in our study, the protein panels are not. Thus, we performed additional tests and
44 experiments to address this important concern.
45
46

47
48 For 188 patients in our study, we compared clinical measurements of CRP from fresh plasma
49 with our MRM-MS measurement from long-term frozen plasma. We calculated robust
50 correlations resistant to outlying points (Rousseeuw PJ., *J Am Stat Assoc* 1984, 79:871 – 880) for
51 the two CRP peptides available in our MRM-MS. The estimated correlation between the clinical
52 and MRM-MS measures was, on average, 0.96 (see Figure A6 below and in the Supplementary
53 Material). Considering the intrinsic differences between these technologies, we interpret these
54 results as a favorable indication of sample stability for this inflammatory protein marker. Results
55
56
57
58
59
60

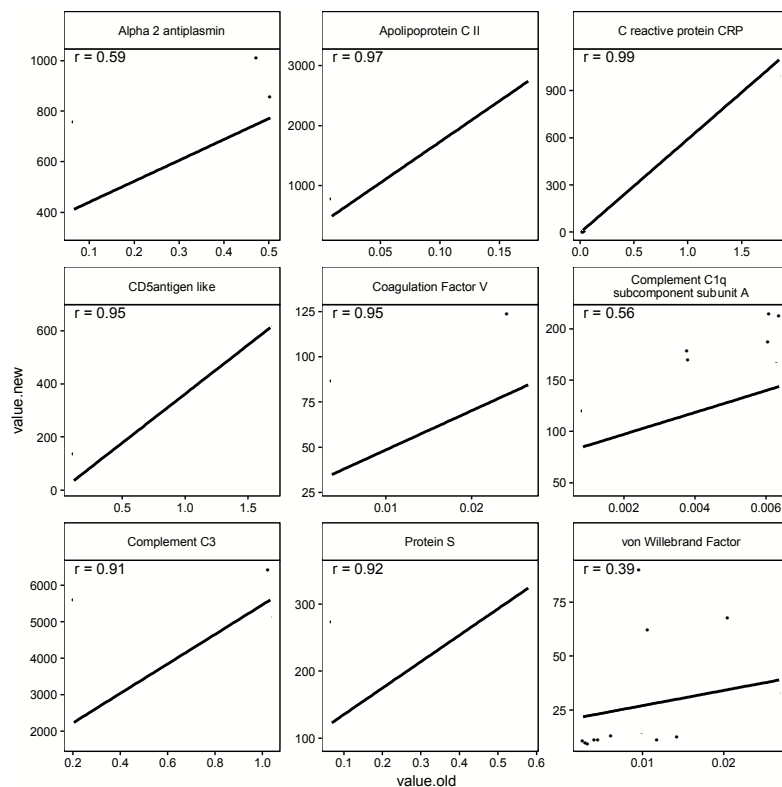
of this analysis are included in the second paragraph of the Discussion (page 19) and in the Supplementary Material. Apolipoprotein B was not measured in our assay.

Figure A6: CRP stability after long-term storage. Correlation between of CRP MRM-MS measurements based on long-term stored samples and ELISA measurements based on fresh plasma samples. MCD robust correlations [1] and sample sizes are shown in the top-left corner.



Unfortunately, there were not samples left from the original study. Thus, we could not measure the proteins in our panel in the same set of samples after 7 extra years of storage. However, in our effort to study this issue, we run an additional experiment using a set of 34 plasma samples measured by Biomarkers in Transplantation (BiT) in January 2012 from a cohort of patients that received a heart transplant. The BiT MRM-MS assay contained peptides from 9 out of the 16 proteins in our panel. For each sample, a separate aliquot of the same EDTA tube was re-analyzed to

study the stability of the additional



study the stability of the additional samples after 6 years of storage.

1
2
3 The robust correlations were above 0.9 for 6 out of the 9 analyzed proteins, about 0.6 for Alpha 2
4 antiplasmin and Complement C1QA, and 0.4 for von Willebrand Factor (see plot above). Since a
5 new set of peptide standards was used in the current analysis, we conclude that most of the
6 proteins analyzed in BiT samples remain stable after a long-term storage. Since samples in this
7 analysis were different from those of our study, we have not included these results in the
8 manuscript. However, we are happy to add these results if the reviewer finds them relevant.
9

10
11 We have also complemented our analysis with a literature review on this important topic. There
12 are several studies that evaluate long term stability of a number of the proteins, including several
13 of the ones investigated in our study. For example, C-reactive protein (CRP) (PMID: 24373927,
14 23516433, 15070788) and Sex hormone-binding globulin (SHBG) (PMID: 17270355, 19290843)
15 have been shown to be stable under long-term storage stability conditions up to 12 years. A study
16 investigating Paraoxonase 1 (PON1) long-term storage at -70°C did not reveal any issues with
17 protein stability (PMID: 22584062). Coagulation factors, such as prothrombin (PMID:
18 23411127), have also been shown to be stable at long-term storage conditions. However, the long
19 term storage stability of some of proteins, such as alpha-2 antiplasmin, apolipoproteins CII and
20 L1, and coagulation factors V and X, in our study, have not been sufficiently investigated in the
21 published scientific literature. We have added this information in the Discussion section as well
22 (page 19).
23
24
25

26 3. *There is much higher prevalence of Diabetes (T1 or T2?) in the CVD group – could the*
27 *model be in part due to diabetes and not CVD?*
28

29 We agree with the reviewer that the prevalence of Diabetes is higher in the CAD group. Thus, the
30 identified panel may be differentiating samples with and without Diabetes, instead of samples
31 with and without CAD. To address this point and remove a potential confounding effect of
32 Diabetes, we used the classifier developed in this study to classify only those samples without
33 Diabetes (n=97 from the discovery set). Results demonstrate that even after removing a potential
34 confounding effect of Diabetes, our classifier can still differentiate between samples with and
35 without CAD (AUCs of 0.76, 0.73, and 0.83 for the clinical, MRM, and ensemble classifiers,
36 respectively). Thus, the performances of the three classifiers remain almost unchanged on this
37 subset of patients without Diabetes.
38
39
40

41 We added a brief comment related to this concern in the Results: Cohort Clinical Characteristics
42 Section (last paragraph of page 15) and in the Supplementary Material.
43

44 4. *There is confusion about what is the response of individual proteins measured. Methods*
45 *indicate 6-point calibration, however all data is presented as log transformed standardized*
46 *relative response (peak area of native/peak area of stable isotope standard) not a*
47 *peptide/protein concentration as would be expected from calibrated data. There is even*
48 *statement, " Calculated Concentration is reported in pmol/μL of plasma or μM" (which is*
49 *not true).*
50
51

52 We understand that our description was confusing since we have not used calculated
53 concentrations but log transformed relative responses in the analysis. We have now clarified that
54 relative responses were used in the analysis and removed unnecessary information from the
55 Materials and Methods section (page 12).
56
57
58
59
60

- 1
2
3
4 5. *It is not clear which “classical risk factors” were used in the reference classifier. Some of*
5 *the are clearly collinear. How were they combined into the classifier since some are*
6 *continuous variables and some binary?*
7

8
9 Predictors are combined using a logistic model. As in a linear regression, predictors of a logistic
10 model can be both continuous and/or binary. Thus, a logistic model can be used to combine both
11 types of explanatory variables in a univariate linear score that explains the logarithm of the
12 subject’s odd of having angiography evidence of CAV. We have included a better description of
13 a logistic model in Materials and Methods-Statistical Analysis (page 13) and in the
14 Supplementary Material.
15

16
17 As expected, some of the explanatory variables in our logistic models are collinear. However, the
18 expected prediction performance of the classifiers is not affected by this problem (a proof can be
19 found in the book by Friedman J, Hastie T, Tibshirani R. The elements of statistical learning.
20 New York, NY, USA, Springer; 2001). A simulation study can be also conducted to prove this
21 point. Collinearity among the explanatory variables of generalized linear models inflates the
22 standard errors of the estimated coefficients, and thus affects their statistical significance in joint
23 models, but does not affect the prediction performance of the estimated classifier. We have also
24 added a reference that explains the difference between significance and prediction (see new
25 reference 20 in the manuscript: Shmueli G. (2010) To explain or to predict? *Statistical science*
26 **25**:289-310).
27
28

- 29
30 6. *There are concerns about the method itself – for majority of proteins the measured peptides*
31 *correlate poorly (median correlation over the 66 detected/measurable proteins is only*
32 *0.82). The selection of the peptide representing protein was made based on highest signal,*
33 *that however, does not guarantee that the selected peptide is true representation of the*
34 *protein abundance and completely different result could be obtained should a different*
35 *peptide was chosen to represent protein abundance. This is demonstrated by selection of a*
36 *peptide for LPA which resides in the kringle IV type 2 domain of LPA and is thus present in*
37 *the sequence in multiple copies and is highly variable due to genetic determination of LPA*
38 *sequence. This is prime example of the need to verify results from targeted proteomics by*
39 *second peptide.*
40
41

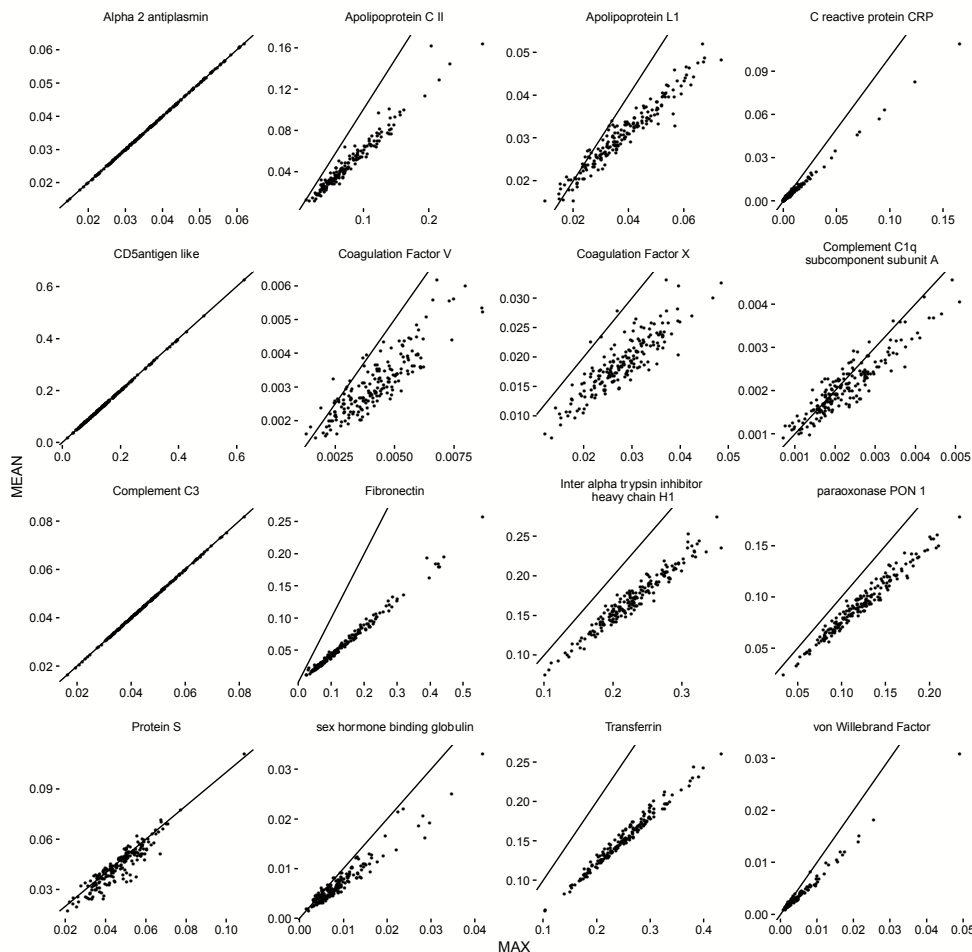
42 We agree with the reviewer that using the peptide with higher signal to represent the protein does
43 not guarantee true representation of protein abundance. While there is not an established
44 consensus on how to integrate the information from different peptides measured for the same
45 protein, it is important to note that our results are robust to using the average level among
46 measured peptides instead. The panel of identified proteins is slightly smaller when MEAN
47 values are used, with 4 proteins dropped: Coagulation Factor V, Coagulation Factor X, Protein S,
48 and Sex Hormone Binding Globulin. However, the performance of the three classifiers (clinical,
49 MRM-MS and ensemble), are very similar regardless of using the peptide with the higher signal
50 (MAX) or using the average abundance (MEAN) to represent each protein (Table A2 added to
51 the Supplementary Material). Similarly, in the test set, the sensitivity and specificity of the
52 ensemble classifier changes from 0.86 and 0.36, respectively, using MAX to 0.78 and 0.45 using
53 MEAN.
54
55
56
57
58
59
60

Table A2: Performance of our 3 classifiers based on clinical risk factors (Clinical), MRM-MS proteins (MRM-MS), and their ensemble (Ensemble) using both the average level among all available peptides per protein (MEAN) and the levels of the peptide with highest values in most samples (MAX).

Cross-validation	AUC			Sensitivity			Specificity		
	Clinical	MRM-MS	Ensemble	Clinical	MRM-MS	Ensemble	Clinical	MRM-MS	Ensemble
MEAN	0.76	0.72	0.83	0.80	0.77	0.87	0.57	0.55	0.61
MAX	0.76	0.75	0.84	0.80	0.77	0.86	0.57	0.60	0.63

To better understand the slight differences observed between both analyses, we compared the MAX *versus* the MEAN abundance of the proteins in the panel in Figure A1 (below and in Supplementary Material).

Figure A1: MAX *versus* the MEAN levels of the proteins in the panel. Black line represents the 45° line where both values are equal.



As expected, there is a strong correlation between both measurements. However, the MEAN value can be highly affected by peptides that cannot be properly measured by the assay shrinking

1
2
3 the measurements of some proteins. Thus, we prefer to keep the analysis based on the MAX
4 measurements. We clearly indicate which peptide was used in the analysis, thus results can be
5 validated with an independent assay or methodology. Given that some researchers may prefer to
6 use the average abundance, we now included these results and discussion in the Supplementary
7 Material and a reference to this analysis in the Materials and Methods section.
8

9
10 We agree with the reviewers that the peptide *GYSTTVTGR* of Apolipoprotein A should have not
11 been included in the analysis. We updated all the results in the manuscript excluding this peptide.
12

13
14 7. *Data for vWF show very unusual pattern with a large group in both healthy and CVD*
15 *groups at identical value of -1 SD – is this because protein was not measurable in these*
16 *people?*
17

18 The peptide used for vWF (ILAGPAGDSNVVK) was not measurable in 28% of the patients.
19 Our original approach uses half of the minimum value observed throughout the assay to impute
20 all non-measurable values (NA) in the data. A second peptide of this protein
21 (IGWPNAPILIQDFETLPR) has even more missing values (41%). Both peptides were missing
22 in 18% of the samples. Importantly, missing values in this protein are not significantly associated
23 with CAD status (Fisher Exact test p-value of 0.1). Thus, we do not interpret the absence or
24 presence of this protein as a sign of CAD.
25

26
27 To reduce the repetition of a single value in multiple samples, we have changed our imputation
28 approach. For any protein containing missing values, if another peptide is available and measured
29 for that protein, then its value is used to impute the missing value. Otherwise, half of the
30 minimum value observed throughout the assay is used to impute. We clarified this in the
31 Materials and Methods section (page 13) and updated all results of the paper accordingly.
32
33

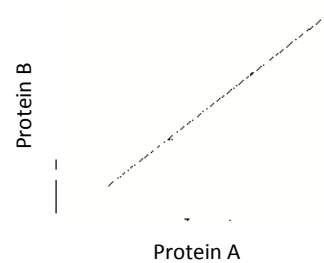
34 8. *There is a lack of detail on how the predictive model was developed. Statement, “Protein*
35 *levels of significant proteins (p-values < 0.05) were combined into a protein classifier*
36 *probability score using logistic regression” is not sufficient – does this mean linear*
37 *combination with CVD status as a binary outcome? Has any optimization been performed to*
38 *test whether some of the proteins are collinear or do not contribute to the prediction? Since*
39 *the p-values used to include proteins into the predictor were not corrected for multiple*
40 *hypothesis testing (the 100 proteins measured or 66 actually used in the analysis) and since*
41 *most of them would not survive this correction, some are likely false positive and likely to*
42 *fall out from the predictor.*
43
44

45 As mentioned previously in point 4, a logistic regression model is a linear combination of the
46 explanatory variables (e.g., protein levels) to explain the logarithm of the odd of success (CAD+
47 in our case). While collinear explanatory variables affect the standard errors of their coefficients
48 and thus their statistical significance, it does not affect the expected prediction performance of
49 the model.
50
51

52 The p-values from the two-sample t-tests in our study were not adjusted for multiple hypothesis
53 testing. While the statistical significance of each marker is relevant to identify potential markers,
54 it is important to note that the prediction performance of individual markers can improve when
55
56
57
58
59
60

all are combined in the logistic model (for a full discussion on this topic see Shmueli G. (2010) To explain or to predict? *Statistical science* **25**:289-310).

For example, the following plot compares the levels of two proteins in case (red points) and control (blue points) samples (toy data). While these proteins cannot discriminate the case and control groups in isolation (see vertical and horizontal overlap of samples from both groups), the diagonal line (linear combination of both proteins) can clearly separate the groups.

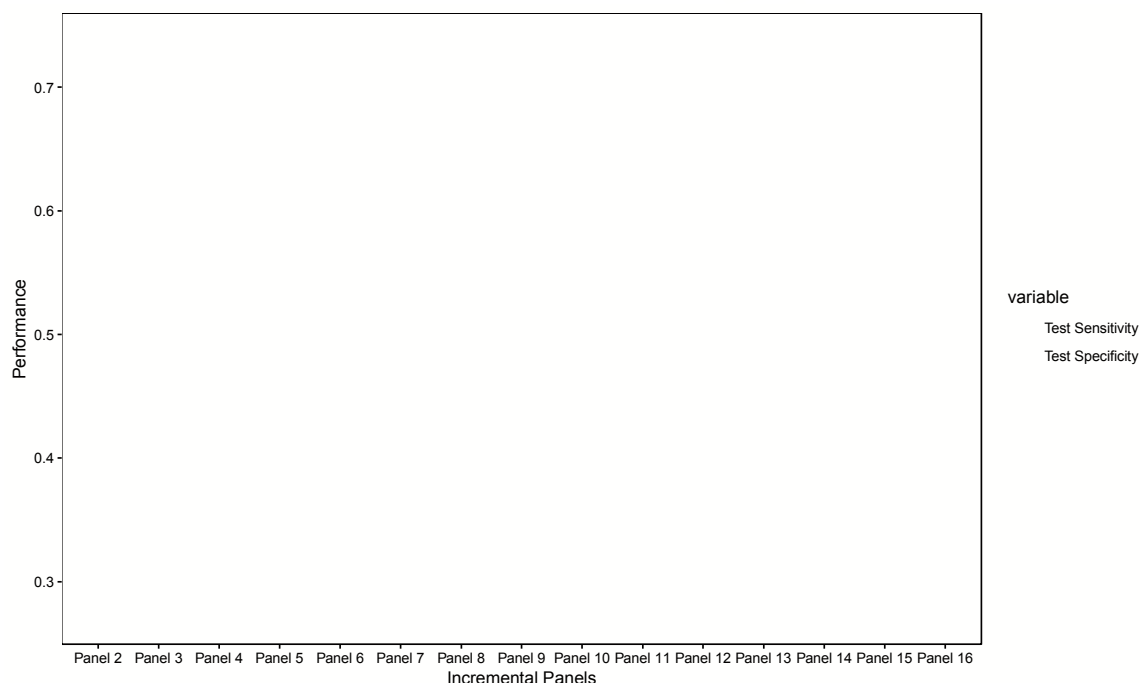


Since our study is focused on the *joint* contribution of proteins and clinical features to predict CAD classes, we have not adjusted the p-values of the univariate tests.

We have now improved our explanation of the logistic regression and the significance criteria used in the Materials and Methods-Statistical Analysis section (page 13-14) and in the Supplementary Material. We have also clarified the use of un-adjusted p-values through the Discussion.

Finally, we have now included a plot illustrating the marginal contribution of each protein to classify samples of the test set. Proteins are added one at a time in order of the univariate p-values (Figure A2 of the Supplementary Material).

Figure A2: Marginal contribution of identified proteins. Sensitivity and specificity values estimated from the test set as proteins are added one at a time to the panel in order of their univariate test p-values (see Table 2 of main text). For example, Panel 2 contains the top-2 proteins (Fibronectin and Protein S) while Panel 16 contains all 16 identified proteins.



- 1
2
3 9. *In Discussion individual proteins should not be discussed as significant unless the p-values*
4 *are still significant after multiple comparison correction. Statements like: " CRP and both*
5 *ApoL1 and ApoCII were significantly elevated in the case group (CAD+) relative to the*
6 *control group (CAD-) in our study (Table 2 and Figure 1, p-values of 0.01, 0.03, and 0.03,*
7 *respectively)" are invalid since with none of these stated p-values would survive correction*
8 *for 66 measurements.*
9

10
11 As recognized by the reviewer, using a stringent univariate filter as that given by individual
12 hypothesis tests with adjusted p-values can control the discovery of false (or weak) individual
13 markers. However, at the same time one may miss the discovery of more subtle but key markers
14 that when combined with others can discriminate patients with CAD from patients without
15 angiographic CAD.
16

17
18 As we explained in the previous point, the primary objective of our study is to predict the status
19 of CAD in patients at risk using a classifier based on multiple proteins. Thus, we have not
20 changed the criteria used to select potential markers. The statistical term "significant" depends on
21 the threshold considered in the analysis, which is stated in the Materials and Methods-Statistical
22 Analysis section (page 13) and in the second paragraph of page 16 in Results. We have further
23 clarified this issue explaining why we have not adjusted the p-values in this study (see Materials
24 and Methods-Statistical Analysis) and adding the word "un-adjusted" when reporting p-values of
25 individual proteins.
26
27

- 28
29 10. *While the method description is detailed some important information is not included – i.e.*
30 *number of concurrent transitions and minimal number of points across the chromatographic*
31 *peak.*
32

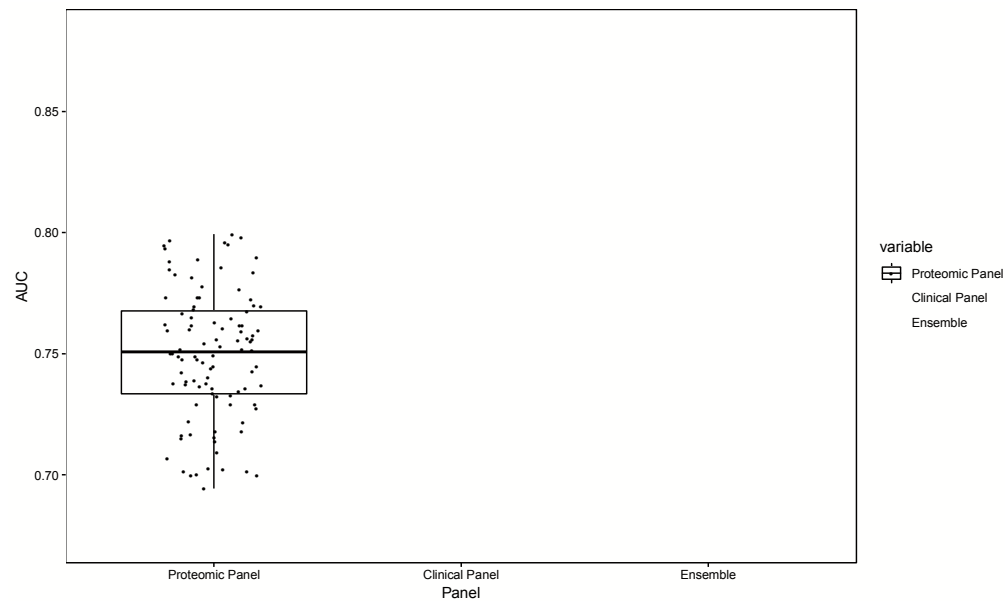
33 We have aimed for a minimum of 10 points across a peak and each peptide was measured with 3
34 transitions. We have now added this information in the Materials and Methods section and have
35 attached a spreadsheet as Supplementary Material to report the qualifier and quantifier
36 transitions.
37

- 38
39 11. *It would be beneficial to report variability encountered in the cross-validation and in the test*
40 *sets. Some of it may be depicted in Figure 3, but the figure is of very low resolution and*
41 *essentially not legible except the axes and curves. Error bars should be presented along both*
42 *axes.*
43

44 As noted by the reviewer, Figure 3 illustrates the receiver operating characteristic curves (ROC)
45 from all cross-validation runs (in light grey). We have improved the resolution of our figures. We
46 hope that the new versions can be easily interpreted. In addition, to better illustrate the variability
47 and distribution of the AUCs computed with cross-validation, we have added boxplots for the
48 AUCs of the 3 classifiers in Figure A5 of the Supplementary Material (and below for your
49 reference).
50
51

52 For the test set, there is no variation in the calculated AUC, sensitivity and specificity since each
53 sample is classified only once and used to compute these performance measures. However, we
54 note that Figure A4 in the Supplementary Material (called A1 in the original version) illustrates
55 the variation of the predicted probabilities for all samples using boxplots.
56
57
58
59
60

Figure A5. Distribution of the area under the receiving operating curve (AUC). Boxplots of the calculated AUCs from 100 runs of 5-fold cross-validations (see Materials and Methods for more details) for each classifier.



1
2
3 **Reviewer: 2**
4

5 **Comments to the Author**

6 *The authors used LC/ESI-MRM-MS measured 242 peptides from 100 plasma proteins and try to*
7 *find potential biomarkers of the coronary artery disease (CAD). The results showed 16 out of the*
8 *100 proteins associated with CAD when comparing the protein concentration in 70 CAD patients*
9 *and 45 control patients. when combined the 16 proteins with clinical measurement, the method*
10 *showed AUC=0.84. This has been confirmed using a different set of subjects (n=87).*
11 *The experiment was well designed and the conclusions seem solid, but a lot of data were not*
12 *shown or not well described, which are necessary to support the results and conclusions.*
13
14

15
16 We thank the reviewer for the deep and useful comments and suggestions, which generated an
17 improved analysis with robust results and a better discussion of our work. Please find a point-to-
18 point response to your comments.
19

20
21 *1. Were all plasma samples collected between 1992-1995?*
22

23 Yes, as indicated in Materials and Methods, the analyzed cohort is comprised of men and women
24 who were referred to two major teaching hospitals in Vancouver for selective coronary
25 angiography between 1992 and 1995. Blood samples were collected in that time frame just
26 before the angiography was performed. We refer the reviewer to our response to point 2 raised by
27 Reviewer 1, which contains more details about our new experiment and analyses of sample
28 stability. We have added a brief discussion of this topic in the Discussion (page 19) and in the
29 Supplementary Material (see Figure A6 in this point 2 of Reviewer 1 and in the Supplementary
30 Material).
31
32

33
34 *2. Transition list is missing. Also need to state which is quantifier and which is qualifier.*
35

36 We have aimed for a minimum of 10 points across a peak and each peptide was measured with 3
37 transitions. We have now added this information in the Materials and Methods section and have
38 attached a spreadsheet as Supplementary Material to report the qualifier and quantifier
39 transitions.
40

41
42 *3. Only measured 127/242 peptides, which passed QC and present in >50% samples. What's the*
43 *reason for the low percentage of usable data?*
44

45 Peptides were excluded from the analysis if they were not measurable in more than 50% of the
46 samples or if they perform poorly in a chromatographic sense (i.e., poor peak shape) leading to
47 high variation during integration and computation of its response. Missing values may be due to
48 peptides present at a level below the LLOQ of the assays. Poor LC performance can be attributed
49 to degradation, incomplete digestion or release, among other reasons.
50

51
52 To obtain robust and reproducible results, we based our analysis on reliable data to demonstrate
53 the potential of using LC/ESI-MRM-MS to discover promising proteins that combined with
54 usual clinical risk factor to enhance risk evaluation of CAD.
55
56
57
58
59
60

1
2
3 We have improved the description of our quality control (QC) criteria in the Materials and
4 Methods section. In addition, in Table A1, we have separated the peptides that were successfully
5 measured by our assay from those that failed our QC criteria as a reference for future studies.
6

7
8 *4. Why the peptides correlation varies? Why choose the most intense peptides? If they are not*
9 *correlate, how to make sure the most intense peptide represents the protein level?*
10

11 We agree with the reviewer that using the peptide with higher signal to represent the protein does
12 not guarantee true representation of protein abundance. While there is not an established
13 consensus on how to integrate the information from different peptides measured for the same
14 protein, it is important to note that our results are robust to using the average level among
15 measured peptides instead. We refer the reviewer to our response to point 6 raised by Reviewer
16 1, which contains full details about our analysis based on average peptide levels. Since the mean
17 can be highly affected by peptides that cannot be properly measured by the assay, we prefer to
18 keep the analysis based on the MAX measurements. We clearly indicate which peptide was used
19 in the analysis, thus results can be validated with an independent assay or methodology. Given
20 that some researchers may prefer to use the average abundance, we now included these results in
21 the Supplementary Material (see Figure A1 and Table A2) and added a reference to this analysis
22 in the Materials and Methods section (page 13).
23
24
25

26 *5. "Calculated Concentration is reported in pmol/ul of plasma" where are these data?*
27

28 We understand that our description was confusing since we have not used calculated
29 concentrations but log transformed relative ratios in the analysis. We have now clarified that
30 relative responses were used in the analysis and removed unnecessary information from the
31 Materials and Methods section (see page 12).
32
33

34 *6. What are the traditional markers currently used in the clinic? Which were used in generating*
35 *the AUC?*
36

37 Traditional blood parameters that are used in the clinic are typically restricted to lipoprotein
38 levels, such as total cholesterol, LDL cholesterol, HDL cholesterol, non-HDL cholesterol,
39 triglycerides, and apolipoprotein B. In some patients, lipoprotein (a) may also be measured, as
40 well as apolipoprotein A-I levels and the apoE genotype. Certain ratios may also be used such as
41 LDL/HDL and ApoB/ApoA1 ratios to further supplement clinical CAD assessment (PMID:
42 24982904). General markers of health including fasting glucose, HbA1c, Creatinine, GFR, ALT,
43 and TSH, may routinely be included in the clinical assessment of CAD. CRP is not routinely
44 used because of its lack of specificity to vascular disease, as it is more of a systematic marker of
45 inflammation.
46
47
48

49 Our clinical classifier includes all the variables that were measured at time of sample collection
50 in these patients and are reported in Table 1 of the manuscript. The aim of the clinical classifier is
51 to combine variables that are relevant in the assessment of CAD into a unique score and compare
52 its performance to that of the MRM-MS model.
53
54
55
56
57
58
59
60

1
2
3 *7. Please deposit raw files and skyline files into public repository.*

4
5 Raw data files have been submitted to PeptideAtlas under the title “Multiplexed LC/ESI-MRM-
6 MS-based Assay for Identification of Coronary Artery Disease Biomarkers in Human Plasma”.
7 We used Mass Hunter to process this data, which is proprietary. Thus, we did not upload any
8 skyline file.
9

10
11 *8. Please discuss how to “reduce a significant burden on healthcare system” by using this*
12 *method? Is it easy to implement this method to clinical labs?*
13

14 All of the biomarkers in Table 1 are routinely measured in the clinic. We show combining these
15 traditional markers with the identified proteins can lead to a better stratification of at-risk
16 patients. While the markers currently used in the clinic are only evident once a certain CAD
17 disease threshold has been reached, the proteomic markers are involved in the early pathogenic
18 processes and can thus help to identify earlier patients with evidence of a more thrombotic and/or
19 pro-inflammatory extracellular environment. By this means, the implementation of this panel in
20 the clinic would reduce significant burden on the health care system. The translation of these
21 markers to the clinic would still require an extensive external validation and a migration of the
22 markers to a clinical assay. We have included this discussion in the second to last paragraph of
23 the manuscript and have softened the conclusive remarks by removing the phrase “*reduce a*
24 *significant burden on healthcare system*”.
25
26
27

28
29 *9. Table 2, add average concentration +/- SD and fold change between 2 groups.*

30 We have reviewed Table 2 to include the additional information requested by the reviewer. More
31 specifically, Table 2 now include average and standard deviations of relative responses of both
32 groups, as well as their fold change. Although the statistical analysis has been performed on log-
33 2 transformed and standardized data (mentioned in the Materials and Methods section), we
34 consider that the raw relative responses summaries for each group can be of interest for the
35 readers and thus reported those in Table 2.
36
37

38
39 *10. Figure 1, vWF showed 2 separate population, why?*
40

41 The peptide used for vWF (ILAGPAGDSNVVK) was not measurable in 28% of the patients.
42 Our original approach uses half of the minimum value observed throughout the assay to impute
43 all non-measurable values (NA) in the data. A second peptide of this protein
44 (IGWPNAPILIQDFETLPR) has even more missing values (41%). Both peptides were missing
45 in 18% of the samples. Importantly, missing values in this protein are not significantly associated
46 with CAD status (Fisher Exact test p-value of 0.1). Thus, we do not interpret the absence or
47 presence of this protein as a sign of CAD.
48
49

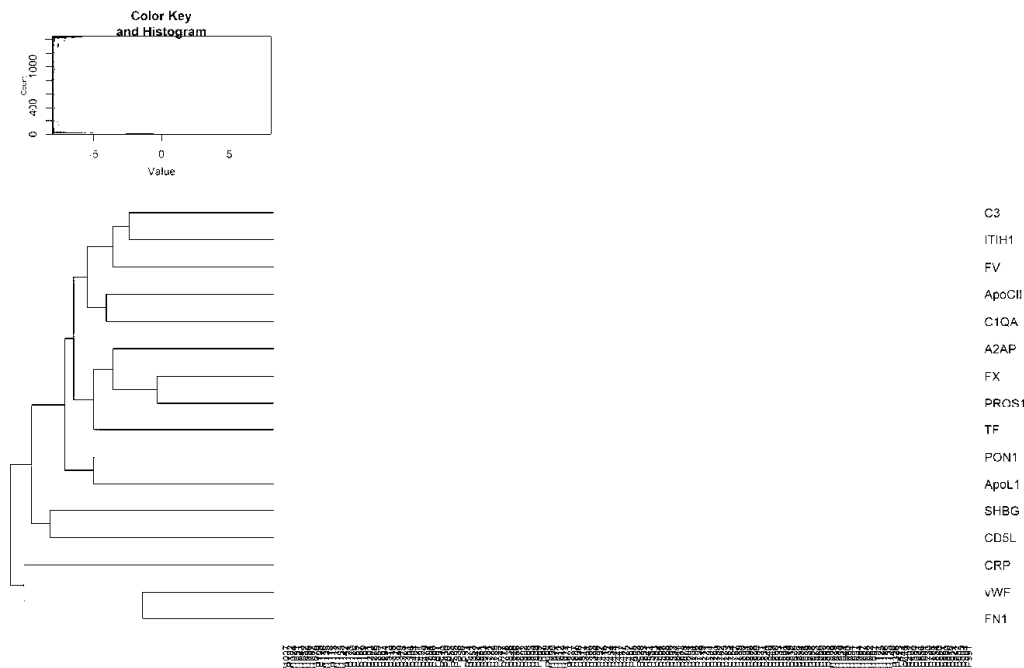
50 To reduce the repetition of a single value in multiple samples, we have changed our imputation
51 approach. For any protein containing missing values, if another peptide is available and measured
52 for that protein, then its value is used to impute the missing value. Otherwise, half of the
53 minimum value observed throughout the assay is used to impute. We clarified this in the
54 Materials and Methods section (page 13) and updated all results of the paper accordingly.
55
56
57
58
59
60

1
2
3 *11. Figure 2, Fibronectin & vWF highly correlate but not with any others, why? It's better to run*
4 *a clustering analysis to see which are clustered together.*
5

6
7 Some studies have shown the co-localization of Fibronectin and vWF in the blood vessel
8 endothelium during cardiovascular injury or disease (PMID: 8533129, 3878055). In fact, both
9 vWF and Fibronectin are known to support the formation of the fibrin clot but it is not known if
10 they work together or independently. Fibronectin can be considered to initiate the first wave of
11 hemostasis and is integrated into fibrin even before platelet accumulation (PMID: 25180602,
12 25180602). As the coagulation cascade continues, coagulation factor XIII (FXIIIa) covalently
13 cross-links vWF to the polymerizing fibrin and fibrin-bound VWF further recruits and activates
14 platelets to support fibrin clot formation (PMID: 25381443).
15

16
17 We have added this information in the Discussion section and included a new figure A3
18 illustrating the cluster analysis suggested by the reviewer in the Supplementary Material.
19

20 **Figure A3.** Cluster analysis of the panel proteins.
21



45 *12. Table A1, What is the unit of the values? If they are the L/H ratio, authors should add ug/ul*
46 *plasma as the SIS are in different concentration so the ratio will have no clinical relevant. Is it*
47 *average (Q1 Q3) or median (Q1 Q3)? how to select the peptides as the representative of each*
48 *protein? For apo(a), GYSTTVTGR is in Kringle repeat region and should not be used as*
49 *representative of protein-level. Add protein accession number*
50
51

52
53 We have now edited Table A1 to clarify and expand the information about the proteins used in
54 our analysis. In this table we report the relative responses of the analyzed peptides, thus there are
55 no units associated. In the reviewed table, we have included the accession numbers of all
56
57
58
59
60

1
2
3 proteins, clarified its caption, and re-order peptides to distinguish better those used in the analysis
4 from those that could not be successfully measured by the developed assay.
5

6 We agree with the reviewers that the peptide *GYSTTVTGR* of Apolipoprotein A should have not
7 been included in the analysis. We updated all the results in the manuscript excluding this peptide.
8
9

10 *13. In abstract, experimental design, "The assay was used to measure 242 stable isotope labeled*
11 *peptide", it also measures the 242 native peptides.*
12

13 We have now added this missing piece of our design. We thank the reviewer for pointing this
14 out.
15

16
17 *14. "Agilent 6490 TQ (QQQ) LC/MS", spell triple quadrupole, LC/MS should be MS*
18

19 We have corrected our spelling and description in the manuscript.
20

21 *15. The LC run time is 36 min. is it possible to reduce the run time in a final method?*
22

23 It is entirely possible to shorten the method so that the proposed methodology becomes more
24 appealing to a clinical lab. However, the data validation in this study was performed using a 36-
25 min method gradient as stated in the Materials and Methods section. Additional validation results
26 would be required to transition into a new gradient. We have added in the sentence in the
27 Discussion indicating that additional validations are required to migrate the identified markers to
28 a clinical assay (page 25).
29
30

31
32 *16. "Endogenous targets whose response is below the lower limit of quantification will have*
33 *calculated concentrations with low accuracies." If it's below LLOQ, is it still worth to report?*
34

35 To control the quality of our data, we have examined the accuracies of the standards points and
36 the limit of quantification of the endogenous targets. Peptides with overall low accuracy and/or
37 values below the limit of quantitation in more than 50% of the samples were eliminated from this
38 analysis. We have improved the description of our quality control (QC) criteria in the Materials
39 and Methods section (pages 12-13). In addition, we have improved the content of Table A1 to
40 describe the measured peptides levels better.
41
42

43 *17. "...enhanced fibrinosis and the incorporation of soluble plasma fibrinogen into the insoluble*
44 *blood clot..." did you see fibrinogen change?*
45

46 This is a good observation. We have not detected significant changes of fibrinogen in this study.
47 We have clarified this statement in the Discussion (page 22).
48
49

50 *18. "...indirectly involved in the pathogenic process, such as in the regulation of cytokines and*
51 *phagocytic cells" Which proteins are involved?*
52

53 To clarify which proteins are involved in the pathogenic process, we have added the following
54 sentences throughout the Discussion section:
55
56
57
58
59
60

1
2
3 Page 21: “Therefore, both ApoL1 and ApoCII interact with lipoproteins, such as HDL, LDL, and
4 VLDL, and are involved in the CAD pathogenic process through mechanisms that are still being
5 researched (Ding, 2014).”
6

7
8 Second paragraph of page 23: “Furthermore, A2AP has a key role in regulating cytokine
9 signaling and phagocytosis activity during inflammation (Kager, 2013), both of which are
10 implicated in the pathogenic process of CAD.”
11

12 Page 24: “Sex hormone-binding globulin (SHBG) is a glycoprotein that is produced and secreted
13 by the liver into the bloodstream where it binds sex steroids, including testosterone and
14 estradiol.” and “Furthermore, the proinflammatory mechanisms that are implicated in CAD
15 pathogenesis through increased cytokine signaling and phagocytosis activity are known to also
16 regulate SHBG activity (Simo, 2015).”
17
18

19 Last paragraph of page 24: “PON1 has important protective effects in cardiovascular diseases,
20 including regulating cytokine signaling and phagocytosis (Martinelli, 2013; Litvinov, 2012).
21 PON1 is also associated with regulating iron metabolism and maintaining proper iron
22 homeostasis (Okuturlar, 2016) and iron dysregulation has been implicated as a risk factor for
23 CAD (Spasojevic-Kalimanovska, 2014; Bagheri, 2013).”
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60