Predictive value of targeted proteomics for coronary plaque morphology in patients with suspected coronary artery disease

Bom, Michiel J; Levin, Evgeni; Driessen, Roel S; Danad, Ibrahim; Van Kuijk, Cornelis C; van Rossum, Albert C; Narula, Jagat; Min, James K; Leipsic, Jonathon A; Belo Pereira, João P

Published in:
EBioMedicine

DOI:
10.1016/j.ebiom.2018.12.033

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2019

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):

Copyright
Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

Take-down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): http://www.rug.nl/research/portal. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.
Research paper

Predictive value of targeted proteomics for coronary plaque morphology in patients with suspected coronary artery disease

Michiel J. Boma\textsuperscript{a,1}, Evgeni Levin\textsuperscript{b,c,1}, Roel S. Driessen\textsuperscript{a}, Ibrahim Danad\textsuperscript{a}, Cornelis C. Van Kuijk\textsuperscript{d}, Albert C. van Rossum\textsuperscript{a}, Jagat Narula\textsuperscript{e}, James K. Min\textsuperscript{f}, Jonathon A. Leipsic\textsuperscript{e}, João P. Belo Pereira\textsuperscript{c}, Charles A. Taylor\textsuperscript{h}, Max Nieuwdorp\textsuperscript{c,j}, Pieter G. Raijmakers\textsuperscript{d}, Wolfgang Koening\textsuperscript{k,j}, Albert K. Groen\textsuperscript{c}, Erik S.G. Stroes\textsuperscript{c}, Paul Knaapen\textsuperscript{a,*}

\textsuperscript{a} Department of Cardiology, Amsterdam UMC, Vrije Universiteit Amsterdam, Amsterdam, the Netherlands
\textsuperscript{b} Horizon BV, Rotterdam, the Netherlands
\textsuperscript{c} Department of Vascular Medicine, Amsterdam UMC, University of Amsterdam, Amsterdam, the Netherlands
\textsuperscript{d} Department of Radiology & Nuclear Medicine, Amsterdam UMC, Vrije Universiteit Amsterdam, Amsterdam, the Netherlands
\textsuperscript{e} Icahn School of Medicine, Mount Sinai Hospital, NY, New York, United States
\textsuperscript{f} Dalio Institute for Cardiovascular Imaging, Weill-Cornell Medical College, NY, New York, United States
\textsuperscript{g} Department of Medicine and Radiology, University of British Columbia, Vancouver, Canada
\textsuperscript{h} Department of Bioengineering, Stanford University, Stanford, CA, United States
\textsuperscript{i} Wallenberg Laboratory, University of Gothenberg, Gothenberg, Sweden
\textsuperscript{j} Department of Internal Medicine, Amsterdam UMC, Vrije Universiteit Amsterdam, Amsterdam, the Netherlands
\textsuperscript{k} Deutsches Herzzentrum München, Technische Universität München, Munich, Germany
\textsuperscript{l} DZHK (German Centre for Cardiovascular Research), Munich Heart Alliance, Munich, Germany

\section*{ABSTRACT}

\textbf{Background:} Risk stratification is crucial to improve tailored therapy in patients with suspected coronary artery disease (CAD). This study investigated the ability of targeted proteomics to predict presence of high-risk plaque or absence of coronary atherosclerosis in patients with suspected CAD, defined by coronary computed tomography angiography (CCTA).

\textbf{Methods:} Patients with suspected CAD (n = 203) underwent CCTA. Plasma levels of 358 proteins were used to generate machine learning models for the presence of CCTA-defined high-risk plaques or complete absence of coronary atherosclerosis. Performance was tested against a clinical model containing generally available clinical characteristics and conventional biomarkers.

\textbf{Findings:} A total of 196 patients with analyzable protein levels (n = 332) was included for analysis. A subset of 35 proteins was identified predicting the presence of high-risk plaques. The developed machine learning model had fair diagnostic performance with an area under the curve (AUC) of 0.79 ± 0.01, outperforming prediction with generally available clinical characteristics (AUC = 0.65 ± 0.04, \(p < 0.05\)). Conversely, a different subset of 34 proteins was predictive for the absence of CAD (AUC = 0.85 ± 0.05), again outperforming prediction with generally available characteristics (AUC = 0.70 ± 0.04, \(p < 0.05\)).

\textbf{Interpretation:} Using machine learning models, trained on targeted proteomics, we defined two complementary protein signatures: one for identification of patients with high-risk plaques and one for identification of patients with absence of CAD. Both biomarker subsets were superior to generally available clinical characteristics and conventional biomarkers in predicting presence of high-risk plaque or absence of coronary atherosclerosis. These promising findings warrant external validation of the value of targeted proteomics to identify cardiovascular risk in outcome studies.

\textbf{Fund:} This study was supported by an unrestricted research grant from HeartFlow Inc. and partly supported by a European Research Area Network on Cardiovascular Diseases (ERA-CVD) grant (ERA CVD JTC2017, OPERATION). Funders had no influence on trial design, data evaluation, and interpretation.

© 2018 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

---

* Corresponding author.
E-mail address: p.knaapen@vumc.nl (P. Knaapen).
1 MJB and EL contributed equally to the manuscript.
Research in context

Evidence before this study

Although current guidelines advice to use risk assessment in patients with stable coronary artery disease, traditional risk models have only moderate prognostic value for future events. Over the last decades, numerous individual inflammatory plasma protein have been linked to the pathophysiology of atherosclerosis in animal models and clinical studies. However, studies on the use of large scale proteomics to identify patients with high-risk coronary plaques as defined by coronary CT angiography are lacking. We performed a PubMed search for articles published up to April 19, 2018, using the following search terms: “biomarker” AND (“coronary CT” OR “CCTA”). This search yielded 25 results. However, none of these studies met our inclusion criteria. Subsequently, we performed a second PubMed search for articles reporting on the use of proteomics to identify patients with stable coronary artery disease at risk for future events. The following search terms were used: “proteomics” AND “coronary artery disease” AND (“events” or “MACE” or “cardiac events”). This search yielded 23 results, of which only one study reported on the use of proteomics to identify patients at risk for coronary events. This study used a large-scale analysis of 1130 plasma proteins to derive and validate a score that was moderately predictive for cardiovascular outcome. Although the concept of using proteomics to identify patients with stable coronary artery disease at risk for future events seems promising, supporting data remains limited.

Added value of this study

Using state-of-the-art machine learning models, trained on targeted plasma proteomics, we defined two complementary protein signatures: one for the identification of patients with high-risk coronary lesions and one for the identification of patients with absence of coronary atherosclerosis on CCTA. Both biomarker subsets were shown to be superior to generally available clinical characteristics and conventional biomarkers in predicting presence of high-risk plaque or absence of coronary atherosclerosis. These promising findings further support the potential of targeted proteomics to identify high risk patients.

Implications of all available evidence

The high diagnostic accuracy of the identified protein signatures for high-risk lesions and the absence of coronary atherosclerosis in the present study, implicates that there could be potential for the use of proteomics for risk stratification in clinical practice. Identification of patients with high-risk lesions can help guide the use of expensive medication to the highest risk group. Conversely, identification of referred patients without coronary atherosclerosis can help avoid overtreatment. Importantly, to confirm the value of the two protein signatures, prospective external validation in outcome studies is warranted. If confirmed, this may herald the introduction of targeted proteomics in the cardiovascular arena.

1. Introduction

Coronary CT angiography (CCTA) is a commonly used non-invasive imaging modality for the diagnostic work-up of patients with suspected coronary artery disease (CAD). In addition to the evaluation of stenosis grade, CCTA allows in-vivo phenotyping of plaque morphology and identification of high-risk plaques, i.e. plaques with a high likelihood of causing an acute coronary event [1–3]. Although current guidelines emphasize the need for adequate risk-assessment in suspected CAD patients [4], traditional risk stratification, using generally available clinical risk factors, plasma lipid levels and other conventional biomarkers (troponin T, NT-proBNP, and CRP) have only modest predictive value for the presence of coronary atherosclerosis and the occurrence of events [5–9]. This underscores the need for novel biomarkers predictive for coronary atherosclerosis. In the search for these biomarkers, several individual inflammatory plasma proteins have been linked to the pathophysiology of atherosclerosis [10]. Recently, the development of proximity extension assays (PEA) has enabled simultaneous measurement of large numbers of proteins using only one microliter of plasma, thereby enabling the use of proteomics in large clinical populations [11]. In parallel, the introduction of machine learning has emerged as a highly effective method for prediction [12]. This study therefore aimed to investigate the ability of a large set of 358 biomarkers to identify either CCTA-derived high-risk coronary lesions or absence of coronary atherosclerosis in patients with suspected CAD using machine learning. Additionally, the added value of these biomarker panels to generally available clinical characteristics and conventional biomarkers was tested.

2. Methods

2.1. Study population

The current report is a sub-study of the PACIFIC trial and details regarding the study design have been described previously [13]. The study population consisted of 208 consecutively selected out-patients with stable new-onset chest pain and suspected CAD. All patients were aged 40 years and above and had intermediate pre-test probability for CAD as defined by Diamond and Forrester criteria. Major exclusion criteria were renal failure (i.e. eGFR <45 mL/min), history of COPD or chronic asthma, a prior history of CAD, atrial fibrillation, and second or third degree AV block. All patients underwent a two day protocol, including CCTA on day 1 and blood sampling on day 2 (within 2 weeks). Both CCTA imaging and laboratory sampling were performed as dictated by the study protocol and independent of clinical status. For the current analysis all patients with CCTA images eligible for plaque analysis and with available laboratory samples were included (n = 203). The study protocol was approved by the Medical Ethics Committee of the VU University Medical Center and written informed consent was obtained.

2.2. Proteomics analysis

Patients were fasting for at least 12 h, resting for at least 10 min and sitting in an upright position before blood sampling was performed. After blood samples were collected, EDTA plasma samples were stored at −80 °C for a mean period of 4 ± 0–8 years before analysis. None of the samples were thawed and refrozen before analysis. After completion of the study, available EDTA plasma samples of the complete cohort were shipped to Olink Proteomics AB (Uppsala, Sweden) for analysis. Using PEA technology, levels of 358 proteins were measured (Supplementary Table 1). The PEA technology has been described previously [11]. In brief, pairs of oligonucleotide-labeled antibody probes bind to their targeted protein, and if the two probes are brought in close proximity the oligonucleotides will hybridize in a pair-wise manner. The addition of a DNA polymerase leads to a proximity-dependent DNA polymerization event, generating a unique PCR target sequence. The resulting DNA sequence is subsequently detected and quantified using a microfluidic real-time PCR instrument (Biomark HD, Fluidigm Corporation, CA, USA). Data is quality controlled and normalized using an internal extension control and an inter-plate control, to adjust for intra- and inter-run variation. The extension control is composed of an
antibody coupled to a unique pair of DNA-tags that serves as a synthetic control that is added to every sample well. It will adjust for technical variation introduced in the extension step and hence reduce intra-assay variability. The final assay read-out is presented in Normalized Protein eXpression (NPX) values, which is an arbitrary unit on a log2-scale where high values correspond to higher protein expressions. Patients with samples that failed quality control were excluded. Furthermore, proteins were excluded from analysis if >20% of individual measurements were below the lower limit of detection. When <20% was below lower limit of detection, missing values were replaced by the limit of detection divided by two. All assay validation data for the proteins in the Cardiovascular II, Cardiovascular III, Cardiometabolic and Inflammation panels (detection limits, intra- and inter-assay precision data, accuracy, etc.) are available on manufacturer’s website (www.okin.com). Across all 92 assays in Cardiovascular II, the mean intra-assay (within-run) and inter-assay (between-run) variations expressed as coefficients of variation are reported to be 9.1% and 11.7%. For assays in the Cardiovascular III panel intra- and inter-assay variation are reported to be 8.1% and 11.4% respectively, for assays in the Inflammation panel 7% and 18% respectively, and for assays in the Cardiometabolic panel 12.9% and 9.5% respectively.

2.3. CCTA acquisition and plaque analysis

Patient preparation, acquisition of CCTA, methodology for plaque analysis, and coronary artery calcium scoring are described in detail in the Supplementary Methods. In short, patients underwent CCTA on a 256-slice CT-scanner (Brilliance iCT, Philips Healthcare, Best, the Netherlands) with prospective ECG-gating (Step & Shoot Cardiac, 256-slice CT-scanner (Brilliance iCT, Philips Healthcare, Best, the Netherlands)) at 75% of the R-R interval. Absence of CAD was

determined by CCTA. Our multivariate analysis showed that patients without CAD had lower systolic blood pressure, male gender, and creatinine and higher sensitivity for hs-Troponin T, NT-proBNP, and CRP levels. Additionally, a third model (referred to as combined model) was computed in which all variables from the clinical model and the predictive biomarkers from the biomarker model were included. This model was tested against the clinical model to establish the incremental predictive value of the identified biomarker panel for high-risk plaque. Subsequently, these three models (clinical, biomarker, and combined) were also computed for the identification of patients with absence of CAD.

2.4. Descriptive statistics

Descriptive statistical analyses were performed using the SPSS software package (version 20.0.0, IBM SPSS Statistics, Armonk, New York). Continuous variables were tested for normal distribution. Normal distributed continuous variables are presented as mean ± SD. Non-normal distributed variables are presented as median with interquartile range. Categorical variables are presented as frequencies with percentages and compared with the chi-square test or Fisher’s exact test were applicable. Variables were compared with the chi-square test for categorical variables and by independent samples t-test for normally distributed continuous variables. Log transformation was applied to NT-proBNP and hs-Troponin T to enable parametric testing.

2.5. Statistical machine learning analysis

We identified two panels of biomarkers that allowed accurate discrimination among 1) patients with high-risk vs. patients without high-risk plaques and 2) patients with presence vs. patients without coronary atherosclerosis, as defined by CCTA. Our multivariate analysis is described in detail in the Supplementary Methods. In brief, we used a combination of deep stacking generalization framework [15] with multiple levels of gradient boosting classifiers [16] to improve prediction accuracy. The strategy employed by our algorithm is to utilize subsampling and model stacking to control over-fitting and improve prediction. The method allows learning of non-linear multivariate relationships among the proteins and is applicable to structured and high-dimensional data.

We conducted a rigorous stability selection procedure [17] to ensure reliability of the biomarker signatures. A randomization test [18] was conducted to evaluate statistical validity of the results. We followed a standard procedure where the outcome variable (e.g., presence of high-risk plaque) was randomly reshuffled while the corresponding protein profiles were kept intact. This was repeated up to 100 times and Receiver-Operating-Characteristics Area-Under-Curve (ROC AUC) scores were computed each time. Predictive value of the biomarker model for the presence of high-risk plaques was tested against a clinical model that comprised generally available clinical characteristics and conventional biomarkers (referred to as clinical model), i.e. age, sex, total cholesterol, HDL cholesterol, diabetes, systolic blood pressure, current smoking, family history, BMI, statin use, eGFR, high-sensitive troponin T, NT-proBNP, and CRP levels. Additionally, a third model (referred to as combined model) was computed in which all variables from the clinical model and the predictive biomarkers from the biomarker model were included. This model was tested against the clinical model to establish the incremental predictive value of the identified biomarker panel for high-risk plaque. Subsequently, these three models (clinical, biomarker, and combined) were also computed for the identification of patients with absence of CAD.

2.6. Data sharing

All individual participant data that underlie the results reported in this article will be made freely available, after de-identification, at Mendeleyn Data (http://dx.doi.org/10.17632/gdfvxxr7f2.1). There are no restrictions to obtaining the data.

3. Results

3.1. Study population

An initial 203 patients with interpretable CCTA images and available laboratory samples was evaluated for inclusion. After laboratory analysis was performed, 7 patients were excluded due to failure to pass the quality control of the proteomics measurements. This resulted in a final study population of 196 patients. The baseline characteristics of these patients are provided in Table 1.

3.2. Coronary CT angiography

CCTA showed one or more coronary lesions with >50% stenosis in 143 (73%) patients. A high-risk coronary lesion was observed in 44 (22%) patients, whereas 26 (13%) patients had complete absence of CAD. An overview of CCTA results is given in Table 2. Baseline characteristics stratified by patient groups, i.e. patients with and without high-risk plaque and patients with and without CAD, are presented in Table 1.

There were some significant differences in baseline characteristics between patients with and without high-risk plaque and between patients with and without CAD (Table 1). Typical chest pain, beta blocker use, and male gender were more frequent and creatinin and hs-Troponin T levels were significantly higher in the high-risk plaque vs non-high-risk plaque group (p < .05 for all). Patients without CAD were younger, more frequently female, less likely to use aspirin and to suffer from hypertension as compared to patients with CAD (p < .05 for all). Furthermore, patients without CAD had lower systolic blood pressure and Framingham Risk Score (p < .05 for all).
3.3. Proteomics analysis

Of all 358 proteins initially included for analysis, 26 had >20% of cases below the lower limit of detection (Supplementary Table 1). These proteins were excluded from further analysis. The total number of proteins included for analysis on plaque morphology prediction was 332.

3.4. Protein signature associated with high-risk coronary lesions

A total number of 35 plasma proteins was predictive for the presence of CT-derived high-risk coronary lesions (Fig. 1 and Supplementary Table 3). The 11 most predictive proteins and the association of protein levels with the presence of high-risk coronary lesions are illustrated in Fig. 1. A complete list of proteins associated with high-risk plaques and the standardized mean protein levels in both groups is provided in the Supplementary Table 3. The machine learning model developed using biomarkers was found to have a high diagnostic value for the presence of high-risk coronary lesions with an AUC of 0.79 ± 0.01 (Fig. 2). Using the same methodology, the clinical model was computed. The relative importance of the variables in the clinical model is shown in Supplementary Table 3. The clinical model had only modest diagnostic accuracy and was inferior to our biomarker panel for the presence of high-risk plaque.

### Table 1
Baseline characteristics.

<table>
<thead>
<tr>
<th>Demographics</th>
<th>Overall (n = 196)</th>
<th>Non-HRP (n = 152)</th>
<th>HRP (n = 44)</th>
<th>p-value</th>
<th>No CAD (n = 26)</th>
<th>CAD (n = 170)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age years</td>
<td>58 ± 8</td>
<td>58 ± 9</td>
<td>59 ± 8</td>
<td>0.35</td>
<td>52 ± 6</td>
<td>59 ± 8</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Male</td>
<td>126 (64%)</td>
<td>90 (59%)</td>
<td>36 (82%)</td>
<td>0.006</td>
<td>12 (46%)</td>
<td>114 (67%)</td>
<td>0.04</td>
</tr>
<tr>
<td>Body mass index</td>
<td>27 ± 4</td>
<td>27 ± 4</td>
<td>27 ± 3</td>
<td>0.85</td>
<td>27 ± 4</td>
<td>27 ± 4</td>
<td>0.63</td>
</tr>
</tbody>
</table>

#### Risk factors - no (%)

- **DM type II**: 30 (15%) vs. 24 (16%) vs. 6 (14%) (p = 0.73 vs. 2% (8%) vs. 28 (17%)) 0.38
- **Hypertension**: 90 (46%) vs. 74 (49%) vs. 16 (36%) (p = 0.15 vs. 7 (27%) vs. 83 (49%)) 0.04
- **Hyperlipidaemia**: 75 (38%) vs. 56 (37%) vs. 19 (43%) (p = 0.45 vs. 8 (31%) vs. 67 (39%)) 0.40
- **Current smoker**: 40 (20%) vs. 28 (18%) vs. 12 (27%) (p = 0.20 vs. 6 (23%) vs. 34 (20%)) 0.72
- **Family history**: 102 (52%) vs. 80 (53%) vs. 22 (50%) (p = 0.76 vs. 7 (27%) vs. 69 (41%)) 0.84

#### Type of chest pain - no (%)

- **Atypical angina**: 76 (39%) vs. 62 (41%) vs. 14 (32%) (p = 0.28 vs. 7 (27%) vs. 69 (41%)) NA
- **Typical angina**: 68 (35%) vs. 44 (29%) vs. 24 (55%) (p = 0.002 vs. 8 (31%) vs. 60 (35%)) NA

#### Laboratory tests

<table>
<thead>
<tr>
<th>TC, mmol/L</th>
<th>LDL-C, mmol/L</th>
<th>HDL-C, mmol/L</th>
<th>hs-Troponin T, ng/L</th>
<th>NT-proBNP, ng/L</th>
<th>Framingham Risk Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.6 ± 1.1</td>
<td>2.5 ± 0.9</td>
<td>1.4 ± 0.5</td>
<td>5.0 ± 4.0</td>
<td>67 ± 40</td>
<td>6.3 ± 3.2</td>
</tr>
</tbody>
</table>

#### Medication use - no (%)

| Statin | 151 (77%) | 112 (74%) | 39 (88%) | 0.04 | 18 (68%) | 133 (78%) | 0.31 |
| Acetylsalicylic acid | 175 (85%) | 133 (88%) | 42 (96%) | 0.17 | 20 (77%) | 155 (91%) | 0.03 |
| Betablocker | 126 (64%) | 90 (59%) | 36 (82%) | 0.006 | 16 (62%) | 110 (65%) | 0.75 |
| ACE-inhibitor/ARB | 73 (37%) | 61 (40%) | 12 (27%) | 0.12 | 6 (23%) | 67 (39%) | 0.11 |

#### Other

| SBP, mm Hg | 143 ± 20 | 143 ± 19 | 144 ± 21 | 0.62 | 135 ± 24 | 144 ± 19 | 0.04 |
| DBP, mm Hg | 82 ± 12 | 82 ± 12 | 84 ± 11 | 0.41 | 79 ± 15 | 83 ± 11 | 0.24 |
| Framingham Risk Score | 6 ± 3 | 6 ± 3 | 6 ± 3 | 0.34 | 4 ± 3 | 4 ± 3 | 0.001 |

### Table 2
CCTA results on a patient-basis.

<table>
<thead>
<tr>
<th>CAC score</th>
<th>Median CAC score</th>
<th>Stenosis - no (%)</th>
<th>No stenosis</th>
<th>0–50% stenosis</th>
<th>50–70% stenosis</th>
<th>&gt;70% stenosis</th>
<th>Plaque analysis - no (%)</th>
<th>No atherosclerotic plaques</th>
<th>Non-calcified plaque</th>
<th>Partially calcified plaque</th>
<th>Calcified plaque</th>
<th>Low attenuation plaque</th>
<th>Positive remodeling</th>
<th>Spotty calcification</th>
<th>Napkin ring sign</th>
<th>High-risk plaque</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-HRP</td>
<td>170 [159–493]</td>
<td>26 (13%)</td>
<td>27 (14%)</td>
<td>53 (27%)</td>
<td>90 (46%)</td>
<td></td>
<td>26 (13%)</td>
<td>78 (40%)</td>
<td>133 (68%)</td>
<td>133 (68%)</td>
<td>57 (29%)</td>
<td>49 (25%)</td>
<td>27 (14%)</td>
<td>20 (10%)</td>
<td>44 (22%)</td>
<td></td>
</tr>
<tr>
<td>HRP</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### Abbreviations:
- CCTA, coronary computed tomography angiography; CAC, coronary artery calcium score.
- Stenosis grade of most severe lesions was used for patient-based analysis.

A total number of 35 plasma proteins was predictive for the presence of CT-derived high-risk coronary lesions (Fig. 1 and Supplementary Table 3). The 11 most predictive proteins and the association of protein levels with the presence of high-risk coronary lesions are illustrated in Fig. 1. A complete list of proteins associated with high-risk plaques and the standardized mean protein levels in both groups is provided in the Supplementary Table 2. The machine learning model developed using biomarkers was found to have a high diagnostic value for the presence of high-risk coronary lesions with an AUC of 0.79 ± 0.01 (Fig. 2). Using the same methodology, the clinical model was computed. The relative importance of the variables in the clinical model is shown in Supplementary Table 3. The clinical model had only modest diagnostic accuracy and was inferior to our biomarker model, with AUC of 0.65 ± 0.04 (p < 0.05). The combined model, comprised of both clinical parameters and the identified predictive biomarkers (AUC of 0.78 ± 0.02) also outperformed the clinical model (p < 0.05). Relative importance of the variables included in the combined model is shown in Supplementary Table 4. All proteomic biomarkers were retained in the combined model, indicating the incremental value of the identified biomarker panel for the presence of high-risk plaque.
3.5. Protein signature associated with absence of coronary artery disease

A different subset of 34 proteins was predictive for the absence of CAD as visualized by CCTA (Fig. 3 and Supplementary Table 5). The 11 most predictive proteins and the association of protein levels with the absence of CAD are illustrated in Fig. 3. An overview of all proteins predictive for the absence of coronary atherosclerosis and the standardized mean protein levels in both groups is provided in Supplementary Table 5. The machine learning biomarker model had a high diagnostic accuracy for the identification of patients with absence of CAD with an AUC of 0.85 ± 0.05 (Fig. 4). The relative importance of the variables in the clinical and combined model are shown in the Supplementary Fig. 1.

Fig. 1. Protein subset predictive for the presence of a high-risk plaque. The importance plot (left panel) illustrates the relative importance of all 35 plasma proteins predictive for the presence of high-risk plaque. The spiderplot (right panel) depicts the 11 most important proteins in our machine learning model that differentiate between the presence (red) and absence of high-risk plaque (green). The axes of the spiderplot represent the standardized mean protein levels (scaled zero-mean unit-variance). Standardized mean levels of MMP12, PLA2G7, TNFRSF10A, TRANCE, REN, TNFRSF13B, MEPE, and CD4 were higher in the high-risk plaque group compared to the non high-risk group. Conversely, TNFRSF10C and SERPINA7 levels were lower in the high-risk group. Abbreviations of protein names are defined in Supplementary Table 1.

Fig. 2. Diagnostic performance of the biomarker model versus the clinical model and the combined model for the presence of high-risk plaque. Receiver-operating characteristics curve with area under the curve (AUC) for the diagnostic performance of the clinical model (left), the biomarker model (middle) and the combined model (right) for the presence of high-risk coronary artery disease. The mean ROC curve for each model is depicted by the blue line. The grey shaded area represents the standard deviation of the curves. The clinical model was outperformed by both the biomarker model ($p < 0.05$) and the combined model ($p < 0.05$).
Tables 6 and 7. Again, the clinical model had only modest diagnostic value ($AUC = 0.70 \pm 0.04$) and was outperformed by the proteomic model ($p < 0.05$). The combined model also had greater diagnostic accuracy than the clinical model with an $AUC$ of $0.83 \pm 0.04$ ($p < 0.05$). All proteomic biomarkers were retained in the combined model, indicating the incremental value of the identified biomarker panel for the absence of CAD.

4. Discussion

This study investigated the ability of a large set of biomarkers to identify patients with either high-risk coronary plaque morphology or absence of coronary atherosclerosis. Results show that a subset of 35 proteins was highly predictive for the presence of high-risk coronary lesions. The diagnostic value of this biomarker signature was significantly...
greater than prediction with generally available clinical characteristics and conventional biomarkers. Additionally, a distinct protein signature of 34 plasma proteins was found to identify patients with absence of coronary atherosclerosis.

4.1. Protein signature for the presence of high-risk plaques

Several studies have investigated the relationship between individual biomarkers and plaque morphology on invasive and non-invasive imaging modalities [9,19,20]. Recent technical advances have enabled the simultaneous measurement of large amounts of proteins and several studies have used these novel techniques to investigate the ability of large biomarker sets to identify patients with angiography-derived significant stenosis [21,22]. LaFramboise et al. studied the predictive ability of 56 proteins for the presence of a coronary stenosis requiring revascularization and found a model that included osteopontin, resistin, MMP7, and IFNγ to be highly predictive [22]. More recently, Ibrahim et al. evaluated the combined predictive value of clinical variables and 109 candidate proteins. The final model comprised several clinical variables and 4 biomarkers (midkine, adiponectin, apolipoprotein C-I, and kidney injury molecule-1) and showed high accuracy for the presence of angiography-derived stenosis. Although interesting from a diagnostic point of view, these studies do not provide any information on the predictive value for the presence of high-risk coronary lesions.

Our study is the first to assess the relationship between a large set of biomarkers and high-risk plaque morphology using high-throughput technology combined with state-of-the-art machine learning techniques. Results show that a protein signature of 35 proteins, many of which have not been linked with atherosclerosis previously, was highly predictive for the presence of high-risk plaques (AUC = 0.79 ± 0.01). A detailed description of the roles of the most important predictive proteins is provided in the supplementary material. Notably, although several predictive proteins identified by LaFramboise et al. and Ibrahim et al. were also tested in our study (osteopontin, kidney injury molecule-1, resistin, IFNγ, and MMP7), none of these markers were included in our biomarker model. However, it is important to note that the predicted outcome in our study (high-risk plaque) was different from these previous studies (angiographic CAD) and therefore the current findings are not necessarily discordant with the aforementioned studies.

For decades, risk prediction in clinical practice has been based on generally available clinical characteristics. However, these clinical characteristics, often combined into risk scores, have only modest predictive value for coronary atherosclerosis and the occurrence of events [5]. More recently, improved cardiovascular risk algorithms have been proposed using additional conventional biomarkers such as high-sensitive troponin, NT-proBNP and CRP [6,8,23]. To test the validity of our findings, our biomarker model was tested against a model which included generally available clinical characteristics and conventional biomarkers. Results show that the biomarker model outperformed this clinical model with AUCs of 0.79 ± 0.01 and 0.65 ± 0.04 respectively. Subsequently, to investigate the incremental value of the biomarker model to clinical risk prediction, the clinical model was tested against a combined model including both proteomic biomarkers and the generally available clinical variables. The combined model (AUC of 0.78 ± 0.02) clearly outperformed the clinical model. Since all proteomic biomarkers were retained in the combined model, the identified biomarker panel had incremental value to generally available clinical characteristics and conventional biomarkers for the prediction of high-risk plaques.

4.2. Protein signature for the absence of coronary artery disease

In addition to the protein signature for high-risk plaques, a different subset of 34 proteins was found to be highly predictive of the CCTA-defined absence of CAD (AUC = 0.85 ± 0.05). The clinical model (AUC = 0.70 ± 0.04) was clearly outperformed by the proteomic biomarker model. Furthermore, the combined model for the absence of CAD (AUC = 0.83 ± 0.04) also more accurately predicted the absence of CAD than the clinical model. Again, all proteomic biomarkers were retained in the combined model, clearly demonstrating the incremental value of the identified biomarker panel to clinical variables for the prediction of the absence of CAD. A detailed description of the roles of the most important predictive proteins is provided in the supplementary material.

4.3. Common pathways in identified protein panels

Interestingly, in the top 11 biomarkers included in our biomarker model for the presence of high-risk plaques (Fig. 1), we identified 7 proteins associated with pro-inflammatory, pro-apoptotic pathways in plaques. Number 1, metalloproteinase-12 (MMP12), a prominent member of the senescence associated secretory phenotype, has been extensively linked with inflammation, plaque stability and atherosclerotic plaque burden [24,25]. Number 2, PLA2G7, also referred to as Lp-PLA2, is an enzyme produced by inflammatory cells, leading to enhanced atherosclerosis by increasing oxidation of LDL particles [26]. Numbers 3, 4, 5, and 8 TNFRSF10A, TNFRSF10C, TRANCE and TNFRS1B, are all involved in the tumor necrosis factor (TNF) signaling pathway and play a role in apoptosis, NFkB activation, and B-cell activation [27]. Number 10, surface CD4, is derived from the T-helper cells, which contribute to the pro-inflammatory cross-talk between T-cells and macrophages among other in the subendothelial compartment [28].

In the top 11 biomarkers for absence of coronary atherosclerosis (Fig. 3), nine are associated with inflammatory pathways. Number 1, GDF-15 (downregulated), a macrophage inhibitory cytokine, is higher in inflammatory states and elevated GDF-15 levels are associated with coronary plaque and occurrence of events [29]. Second, CCL24 (downregulated) contributes to recruitment of neutrophils and macrophages and inhibition of CCL24 is shown to attenuate inflammatory activity in pro-inflammatory disorders [30,31]. Third, chitinase-1 (CHIT1), an enzyme secreted by activated macrophages, has been shown to be abundant in atherosclerotic plaques [32]. Interestingly, MMP12 and PLA2G7 were also present in the panel predictive for high-risk plaques. Closer exploration of the raw data revealed these proteins to be upregulated in the high-risk group but downregulated in the absence of CAD (Figs. 1 and 3). In total, 6 proteins (i.e. MMP12, PLA2G7, renin, PRSS27, SLAMF1, and TNFRSF9) were present in both protein signatures. All six proteins show opposite regulation in the high-risk plaque versus the absence of CAD (Supplementary Tables 2 and 5), confirming their potential biological involve in plaque formation. Number 9 and 10, osteoprotegerin (OPG) and TNFRSF9, are both members of the TNF receptor family. OPG (downregulated) is reported to harbor protective effect by serving as decoy receptor for receptor-activator of nuclear factor kappa-B ligand and as inhibitor of vascular calcification [27,33]. However, several studies have associated elevated OPG levels with cardiovascular events [34]. TNFRSF9 (downregulated) plays a key role in activating pro-inflammatory T-helper cells and is implicated in the progression of atherosclerosis [35].

Upon integration of the ‘high-risk plaque’ and ‘absence of CAD’ signatures, a concept emerges of a systemically quantifiable hyperactivity of pro-inflammatory signaling cascades involved in T-cell, macrophage, apoptotic activity, and cellular recruitment in case of the presence of high-risk plaques. Conversely, there is hypo-activity of these signaling pathways in case of absence of coronary atherosclerosis.

4.4. Clinical implications and future perspectives

Despite recent advances in lipid-lowering therapies, patients with CAD remain at substantial risk for coronary events [36]. Elevated levels of biomarkers of vascular inflammation such as CRP have been proposed to indicate the extent of residual risk in these patients. The promise of this concept was recently confirmed by the CANTOS trial which showed...
that in patients with elevated CRP levels, anti-inflammatory therapy targeting interleukin-1β reduced the rate of cardiovascular events independent of lipid-level lowering [37]. These findings may be seen as a first step towards patient-tailored management in CAD patients, in which phenotyping of both residual lipid- and inflammatory-driven risk help guide therapeutic choices [38]. Given the non-specific nature of CRP, there is a need for more specific markers of the inflammatory processes in the arterial wall. Technical advances in proteomic analyses paralleled by advances in machine learning have paved the way for the introduction of proteomics. Given the relatively low costs compared to CCTA and routine laboratory testing of single biomarkers, and the rapid analysis of proteomic samples (within 24 h), targeted proteomics offers the opportunity to be highly cost-effective. Ganz et al. recently demonstrated the ability of targeted proteomics, measured using modified aptamers, to identify patients at risk for cardiac events [39]. Given the limited patient population in our study cohort, validation with clinical events was not possible. Alternatively, we opted to investigate the predictive value of targeted proteomics for CCTA-defined high-risk coronary lesions and absence of coronary atherosclerosis. Since multiple studies have confirmed the link between CT-derived high-risk plaques and cardiovascular events [1,2] and conversely between the absence of CAD and event-free survival [40,41], prognostic implication of both identified protein signatures seems plausible. Of course, our results must be seen as hypothesis-generating and validation in external cohorts with clinical events is necessary for clinical application. Hypothetically, identification of high-risk patients could help guide the use of expensive medication to the highest risk groups, with the promise of shedding light on the predominant risk factors in these patients (lipid, inflammation, thrombosis). Conversely, identification of patients without coronary atherosclerosis could help avoid overtreatment. Importantly, to confirm the prognostic value of the two protein signatures, prospective validation in outcome studies is warranted. If confirmed, this may herald the introduction of targeted proteomics in the cardiovascular arena.

4.5. Limitations

Several aspects may need closer attention. First, the current report is a substudy of the PACIFIC trial and although blood samples were prospectively drawn and stored for analysis, the current report has a limited sample size and the analysis should be considered retrospective in nature. Second, external validation of our findings in a separate validation cohort was not possible due to the relatively small sample size. To minimize the risk of overfitting, state-of-the-art machine learning modelling was performed best suited to the nature of the data and the limited sample size. A rigorous stability selection procedure and specialized regularization strategy were used to ensure reliability of our findings. Subsequently, biomarker signatures were internally validated in the current cohort to avoid over-fitting, by using a 10-fold stratified cross-validation over the training partition of the data (80%) while the remaining 20% was used as the testing dataset. For increased confidence, this procedure was repeated multiple times on a reshuffled dataset. Although we believe these techniques corroborate the validity of our findings, external validation in larger cohorts is mandatory to confirm the predictive value of the identified biomarker subsets. Similarly, validation with coronary events was not possible in this cohort due to the relatively small sample size. Instead, the presence of high-risk plaque was used as a surrogate for coronary events. It must however be noted that although high-risk plaque has been extensively associated with adverse outcome, not all high-risk plaques cause events [1,2,14]. Validation with coronary events in an external cohort is therefore crucial to validate our findings. Third, our study population consisted of symptomatic patients with intermediate risk of CAD. Therefore, our prediction model may be less suitable for subjects in a low-risk, primary prevention cohort. Fourth, our predictive methodology focused on the performance of the developed machine learning model involving a joint panel of selected biomarkers which together, as a group, lead to reliable prediction. This is a crucial difference in comparison with univariate models that evaluate the up- or downregulation of single proteins. Contribution of all proteins included in the joint panel is needed to obtain a reliable prediction model. Our study therefore does not provide definite information on the up- or downregulation of single proteins. Last, since some of the included proteins have not been implicated in the pathophysiology of atherosclerosis before, further studies are warranted to elucidate their role in atherogenesis.

5. Conclusions

Using statistical machine learning models, trained on targeted plasma proteomics, we defined two complementary protein signatures: one for the identification of patients with high-risk coronary lesions and one for the identification of patients with absence of coronary atherosclerosis. Both biomarker subsets were shown to be superior to generally available clinical characteristics and conventional biomarkers in predicting presence or absence of (high-risk) coronary atherosclerosis. These promising findings warrant external validation of the value of targeted proteomics to identify cardiovascular risk in cardiovascular outcome studies.

Funding sources

This study was supported by an unrestricted research grant from HeartFlow Inc. and partly supported by a European Research Area Network on Cardiovascular Diseases (ERA-CVD) grant (ERA-CVD-JTC2017, OPERATION). Funders had no influence on trial design, data evaluation, and interpretation.

Declaration of interests

Drs. Leipsic has core laboratory contracts with Edwards Lifesciences for which he receives no direct compensation; and has served as a consultant for and received stock options from Circle CVI and HeartFlow. Drs. Min reports serving as a consultant to HeartFlow Inc. and Abbott Vascular, serving on the scientific advisory board of Aritena, and holding an equity interest in MDXdx. Dr. Taylor is an employee and shareholder of HeartFlow, Inc. Dr. Koenig has received lecture and consultancy fees from Novartis, Amgen, and AstraZeneca; has received lecture fees from GlaxoSmithKline, The Medicines Company, Pfizer, and Merck Sharpe & Dohme; and has received research grants from Roche Diagnostics, Abbott, Singulex, and Beckmann. Dr. Nieuwdorp is supported by a Le Ducq consortium grant 17CVD001 and Novo Nordisk Foundation GUT-MMM grant as well as by a ZONMW-VIDI grant 2013 (816.146.327) and CVON Young Talent grant 2012. Dr. Levin is founder of HorAlzon B.V. The other authors have no conflicts to report.

Author contributions: All authors contributed equally in the conception and design of the study. M.B., E.L., R.D., I.D., J.L., J.M., A.G., E.S. and P.K. drafted the manuscript. All authors read, critically revised and approved the manuscript.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ebiom.2018.12.033.

References
