DOI: 10.1161/CIRCGENETICS.113.000490

Simultaneous Consideration of Multiple Candidate Protein Biomarkers for Long-Term Risk for Cardiovascular Events

Running title: *Halim et al.; Multiple Proteins Associated with Death/MI & Death*

Sharif A. Halim, MD, MHS[1,2]; Megan L. Neely, PhD[2]; Karen S. Pieper, MS[2];

Svati H. Shah, MD, MHS[1-3]; William E. Kraus, MD[1]; Elizabeth R. Hauser, PhD[3];

Robert M. Califf, MD[1,4]; Christopher B. Granger, MD[1,2]; L. Kristin Newby, MD, MHS[1,2]

[1]Division of Cardiology, [2]Duke Clinical Research Institute, [3]Duke Center for Human Genetics, [4]Duke Translational Medicine Institute, Duke University Medical Center, Durham, NC

Correspondence:

L. Kristin Newby, MD, MHS

Duke Clinical Research Institute

P.O. Box 17969

Durham, NC 27715

Tel: (919) 668-8805

Fax: (919) 668-7056

E-mail: kristin.newby@duke.edu

Journal Subject Codes: [135] Risk factors, [144] Other arteriosclerosis, [4] Acute myocardial infarction

DOI: 10.1161/CIRCGENETICS.113.000490

Abstract:

Background - Although individual protein biomarkers are associated with cardiovascular risk, rarely have multiple proteins been considered simultaneously to identify which set of proteins best predicts risk.

Methods and Results - In a nested case-control study of 273 death/myocardial infarction (MI) cases and 273 age- (within 10 years), sex-, and race-matched and event-free controls from among 2023 consecutive patients (median follow-up 2.5 years) with suspected coronary disease, plasma levels of 53 previously reported biomarkers of cardiovascular risk were determined in a core laboratory. Three penalized logistic regression models were fit using the elastic net to identify panels of proteins independently associated with death/MI: proteins alone (Model 1); proteins in a model constrained to retain clinical variables (Model 2); and proteins and clinical variables available for selection (Model 3). Model 1 identified 6 biomarkers strongly associated with death/MI: ICAM-1, MMP-3, NT-proBNP, IL-6, sCD40L, and IGFBP2. In Model 2, only sCD40L remained strongly associated with death/MI when all clinical risk predictors were retained. Model 3 identified a set of 6 biomarkers (ICAM-1, MMP-3, NT-proBNP, IL-6, sCD40L, and IGFBP2) and 5 clinical variables (age, red-cell distribution width, diabetes, hemoglobin, and New York Heart Association class) strongly associated with death/MI.

Conclusions - Simultaneously assessing the association between multiple putative protein biomarkers of cardiovascular risk and clinical outcomes is useful in identifying relevant biomarker panels for further assessment.

Key words: protein, prognosis, mortality, myocardial infarction

DOI: 10.1161/CIRCGENETICS.113.000490

Despite advances in identification, risk stratification, and treatment of individuals with coronary artery disease over the past 30 years, there remain limitations to identifying patients at the highest risk of adverse events who would potentially benefit most from more aggressive therapies. Data from large clinical trials provide an estimate of the average response to a particular intervention. Stratified medicine focuses on understanding the unique clinical and biological characteristics of smaller groups of individuals in an attempt to better assess risk and predict response to treatment. Evolving genomic, proteomic, and metabolic profiling may facilitate understanding of disease processes and provide novel blood-based biomarkers that will further refine our ability to not only stratify risk but also tailor therapies according to unique molecular profiles. Understanding the interplay of biomarkers from different pathways is paramount to moving the field forward.

Despite hundreds of reports associating individual protein biomarkers with risk for death or myocardial infarction (MI), few studies have considered more than a few biomarkers simultaneously or in the context of well-defined clinical risk modeling. As a first step to addressing this need, we used the MURDOCK (Measurement to Understand the Reclassification of Disease of Cabarrus and Kannapolis) Horizon 1 Cardiovascular Disease (H1 CV) Study to perform a nested case-control study. We analyzed stored plasma samples to determine levels of 53 previously identified putative protein biomarkers of risk for death and death/MI that reflect multiple pathways relevant to cardiovascular pathophysiology (i.e., inflammation and atherosclerosis, myocardial necrosis, thrombosis, endothelial dysfunction and extracellular matrix remodeling, hemodynamic stress, and metabolism); from these, we aimed to identify smaller panels of biomarkers independently associated with clinical outcomes.

DOI: 10.1161/CIRCGENETICS.113.000490

Methods

Patient Population

The MURDOCK H1 CV study has been previously described.[1] The overall study cohort consists of 6447 patients who underwent coronary angiography for known or suspected coronary artery disease and were enrolled in the Duke CATHeterization GENetics (CATHGEN) biorepository (http://cathgen.duhs.duke.edu) from January 1, 2001 to November 14, 2007. All patients were free of pulmonary hypertension, severe lung disease, advanced heart failure (defined as New York Heart Association [NYHA] class IV and systolic dysfunction [ejection fraction <35%]), congenital heart disease, and prior solid organ transplant. From within this cohort, we identified a molecular profiling cohort of 2023 sequential patients (median duration of follow-up 2.5 years) who all provided plasma, DNA, and RNA at the enrollment cardiac catheterization. Within this consecutive cohort, a nested case-control cohort was identified, consisting of all 273 cases with death/MI occurring any time after the index cardiac catheterization and 273 controls without events matched for age (within 10 years), sex (exact match), and race (exact match). Patients were matched using the %match SAS macro,[2] which implements the optimal matching algorithm.[3] The optimal algorithm sorts cases and controls, then finds all pairs that satisfy the specified distance measures, and then selects the set of pairs that minimizes the total distance between all pairs.

Plasma used for protein analysis was prepared from EDTA tubes that were collected after insertion of the arterial sheath for the catheterization. Samples were centrifuged at 4°C at 1500*g* for 10 minutes within 30 minutes of collection, separated into 0.5 mL aliquots, and stored at ¯80°C. Clinical data were provided from the Duke Databank for Cardiovascular Disease (DDCD), which archived clinical and procedural data and longitudinal follow-up information for

all patients in the cohort. Supplemental clinical data that were not contained in the DDCD were obtained from the Duke Decision Support Repository or from direct review of medical records. Endpoint MI was defined as creatine kinase-MB or troponin I or T levels greater than the upper limit of normal in patients with chest pain, cardiac arrest, or other symptoms suggestive of cardiac ischemia. Death was confirmed through the Social Security Death Index and National Death Index as a part of standard clinical follow-up in the DDCD.

The CATHGEN biorepository is approved by the Duke University Institutional Review Board (IRB), and all participants provided written informed consent. Use of CATHGEN samples and Duke clinical data for the MURDOCK H1 CV study was approved by the Duke IRB with a waiver of informed consent and Health Insurance Portability and Accountability Act authorization.

Assay Selection and Proteins Tested

We selected 53 proteins for analysis based on previously published evidence that suggested their association with risk of death or a composite of death/MI among patients with suspected or confirmed cardiovascular disease (CVD) or with risk factors for CVD. We also included potential novel biomarkers for which commercial assays were available on 1 of 2 multiplexing platforms: Meso Scale Discovery (Rockville, MD) and Luminex (Austin, TX) for protein assays unavailable through Meso Scale Discovery. All assays on the Meso Scale Discovery and Luminex platforms were performed at the David H. Murdock Research Institute Core Laboratory in Kannapolis, NC. Lipoprotein-associated phospholipase A2 (Lp-PLA2) was assayed using the PLAC® Test and the colorimetric activity method (CAM) at diaDexus, Inc. (South San Francisco, CA). The full list of proteins analyzed and the analytical characteristics of the assays are provided in the Supplemental Material. All sample analyses were conducted blinded to case-

control status.

Statistical Methods

Baseline characteristics were summarized as medians with interquartile ranges for continuous variables and counts with proportions for categorical variables. Continuous variables were compared using the Kruskal-Wallis test, and categorical variables were compared using Pearson's chi-square test. Biomarker levels were summarized as median concentrations with interquartile ranges. To identify a panel of biomarkers associated with death/MI (primary) or death (secondary) from among the 53 putative protein biomarkers, we used penalized logistic regression to perform variable selection. Prior to regression modeling, we assessed the correlation among the putative biomarkers using the Spearman rank test. In addition, the linearity of the relationship between each biomarker and outcomes (death/MI and death) was assessed using a chi-square goodness-of-fit test. Because nonlinearity was suspected, the test for linearity was performed for the raw biomarker data as well as for log-base-2–transformed data. For death/MI, 8 biomarkers had significant nonlinear relationships after log transformation compared with 28 biomarkers before transformation ($P<0.05$; Supplemental Material). For death, 7 biomarkers had significant nonlinear relationships after log transformation compared with 24 biomarkers before transformation ($P<0.05$; Supplemental Material). Based on these findings, we performed variable selection on the log-transformed values using the penalized regression method known as the elastic net.[4] The elastic net places a penalty on the size of the estimate coefficients in the likelihood function being optimized, shrinking the estimated coefficients of non-important predictors to zero. As such, the approach is able to perform coefficient estimation and variable selection simultaneously.

We fit 3 models for each outcome (death/MI and death) using the elastic net method with

DOI: 10.1161/CIRCGENETICS.113.000490

5-fold cross validation. Model 1 evaluated the 53 putative proteins alone to identify a subset of protein biomarkers associated with death/MI and death; Model 2 identified a subset of proteins associated with outcomes, adjusting for clinical covariates identified in prior modeling as predictors of death/MI and death in the overall MURDOCK H1 CV cohort; and Model 3 allowed for variable selection among both candidate proteins and the previously identified clinical covariates. Clinical variables adjusted for in death models included age, sex, weight, blood pressure, heart rate, smoking history, diabetes, presence of chest pain at presentation, NYHA class, ejection fraction, atrial fibrillation, left bundle branch block, left ventricular hypertrophy, corrected QT interval, red cell distribution width (RDW), serum sodium, blood urea nitrogen, creatinine, hemoglobin, white blood cell count, Duke coronary artery disease index, and Charlson comorbidity index. Clinical variables adjusted for in death/MI models included all the previously mentioned variables except serum sodium and left ventricular hypertrophy.

Fitting a penalized regression model with cross validation on the same data more than once might produce different lists of significant predictors; to overcome this, we generated 500 bootstrap samples of the data for each response and found the proportion of times each candidate variable was included in the model. Evidence of association was defined as strong (selection in ≥85% of the samples) or moderate (≥70% but <85% of the samples). We assessed model discrimination for all models with C-indices and constructed a 95% confidence interval for each index to identify statistically significant differences between the models. Reported odds ratios (ORs) were calculated using logistic regression models of the variables selected from the elastic net. Therefore, they should be used to provide insight into the direction of the association and not the magnitude of effect.

Sensitivity Analysis

A second analysis was performed using the group LASSO method after transforming the biomarker data using piecewise linear splines, thus resolving all instances of nonlinearity.[5] Differences between the elastic net and the group LASSO are detailed in the Supplemental Material.

Results

Patient Characteristics and Plasma Biomarker Levels

Baseline clinical characteristics are displayed in Table 1, and concentrations of the proteins assayed are displayed in Table 2. Pregnancy-associated plasma protein A (PAPP-A) and interleukin-1 alpha (IL-1α) were removed from the analysis because plasma levels were not detectable in our population.

Association of Biomarkers with Clinical Outcomes

Tables 3 (death/MI) and 4 (death) show the proteins with strong and moderate evidence of association with outcome in Models 1 and 2 as assessed by the percentages of bootstrapped samples in which they were selected. ORs for each biomarker are provided as a measure of the direction of association. Table 5 displays Model 3 results for both death/MI and death.

Death/MI

Model 1: biomarkers alone

Intercellular adhesion molecule-1 (ICAM-1), matrix metalloproteinase-3 (MMP-3), soluble CD40 ligand (sCD40L), interleukin-6 (IL-6), insulin-like growth factor binding protein-2 (IGFBP2), and N-terminal pro-B-type natriuretic peptide (NT-proBNP) demonstrated strong evidence of association. No biomarker showed moderate association with death/MI.

Model 2: biomarkers in models constrained to retain all clinical covariates

Only sCD40L demonstrated strong association with death/MI after adjustment for clinical covariates. Biomarkers with moderate association after adjustment for clinical variables included ICAM-1, MMP-3, IL-6, and IGFBP2.

Model 3: candidate proteins and clinical variables all allowed for selection

MMP-3, sCD40L, ICAM-1, IL-6, NT-proBNP, IGFBP2, NYHA class, RDW, hemoglobin, diabetes, and age were strongly associated with death/MI. Additional variables that were moderately associated with death/MI included soluble Fms-like tyrosine kinase-1 (sFlt-1), placental growth factor (PlGF), IL-1 receptor antagonist (IL-1RA), smoking status, serum creatinine, systolic blood pressure, ejection fraction, weight, and Duke coronary artery disease index.

Death

Model 1: biomarkers alone

Five biomarkers were strongly associated with death: MMP-3, NT-proBNP, IGFBP2, D-Dimer, and IL-6. An additional 4 biomarkers were moderately associated with death: vascular cell adhesion molecule-1 (VCAM-1), ICAM-1, sCD40L, and growth differentiation factor-15 (GDF-15).

Model 2: biomarkers in models constrained to retain all clinical covariates

After adjusting for clinical covariates, MMP-3, IL-6, and sCD40L had moderate evidence of association with mortality, but no biomarker demonstrated strong association.

Model 3: candidate proteins and clinical variables all allowed for selection

NT-proBNP, MMP-3, IL-6, NYHA class, RDW, serum creatinine, and age had strong evidence of association with mortality. Variables demonstrating moderate evidence of association with

death included VCAM-1, ICAM-1, PlGF, sCD40L, D-Dimer, IGFBP2, serum sodium, ejection fraction, and diastolic blood pressure.

Model discrimination

C-indices with 95% confidence intervals for a model with only clinical variables (Model 0) and for the 3 biomarker-related models are displayed in Table 6. The increments in C-index were not statistically significant.

Sensitivity Analysis

Like the elastic net models, the group LASSO models assessed for strong and moderate associations between biomarker levels and death/MI and death outcomes. Overall, the selected biomarkers were the same as those in the elastic net with the exception of RANTES (regulated on activation normal T cell expressed and secreted), platelet-derived growth factor (PDGF) AB/BB, and interleukin-18 (IL-18). RANTES and PDGF AB/BB levels were strongly correlated with one another and with sCD40L levels. Biomarkers with moderate or strong associations with death/MI and death using the group LASSO are provided in the Supplemental Material.

Discussion

In this study of 546 patients undergoing angiography for known or suspected coronary disease, we employed penalized logistic regression to simultaneously assess the relationships of multiple putative, highly correlated protein biomarkers with death/MI and death during a median of 2.5 years of follow-up. Using this approach and a set of 53 previously identified putative biomarkers, we identified smaller sets of biomarkers that were independently associated with clinical events in the context of all other putative markers. Furthermore, we identified several biomarkers that were associated with events after adjusting for known baseline clinical covariates. Importantly, proteins representing different mechanistic pathways were selected into the models, reinforcing

the potential relevance of this method for selecting biomarker panels for future study and development as potential clinical or research tools.

Our results highlight several important considerations. Of the many prior studies demonstrating individual biomarker associations with adverse outcomes, few have assessed the associations of these biomarkers with outcomes simultaneously in the context of one another. In addition to testing multiple candidate biomarkers simultaneously, we also assessed the independent association of multiple biomarkers simultaneously in the context of clinical variables associated with death/MI or death. This strategy is important because biomarker-based risk stratification should contribute information beyond readily available clinical data. Furthermore, as many of the proteins assayed were correlated with one another, our statistical method allowed us to select important covariates as a group rather than potentially arbitrarily selecting 1 marker from among the cluster of correlated biomarkers.

Role of Multiple Pathways in Adverse Cardiac Outcomes

Our analyses identified candidate proteins that were strongly associated with death/MI or death that represented inflammation and atherosclerosis, vascular/endothelial dysfunction and extracellular matrix remodeling, hemodynamic stress, metabolism, and thrombosis pathways. These observations highlight the importance of understanding the roles and interplay of multiple biological pathways in the development of CVD and ischemic events. Importantly, for developing candidates for further evaluation for clinical utility, we found that some proteins previously shown to play a key role in the pathogenesis of CVD may not be as important when evaluated in the context of proteins from other pathways, or even may not be the most important protein in their own pathways. For example, in our analysis, IL-6 (reflecting the inflammatory pathway) was consistently associated with death/MI and death in all 3 models, and VCAM-1 and

ICAM-1 were also strongly associated with events. However, several biomarkers of inflammation associated with clinical events in other studies were not associated with clinical events in our multiplexed assessment, including high-sensitivity C-reactive protein (hsCRP), which is downstream of IL-6 and has been associated with recurrent events and benefits from aggressive primary prevention.[6-9] This observation may reflect that among a group of correlated markers reflecting the inflammatory pathway, the markers selected were more strongly associated with events than with hsCRP. Similarly, markers of myonecrosis were not consistently associated with outcomes in our analyses, though a limitation is that we did not evaluate newer highly sensitive troponin assays, which are associated with death or future heart failure even in apparently normal individuals.[10,11] Furthermore, given the performance characteristics of the troponin assay used, the lack of association with clinical outcomes compared with current contemporary sensitive assays in other studies may be related to the assay's lack of low-end accuracy. An alternative explanation may be related to the time horizon over which we identified event cases. Our median follow up was 2.5 years, but other studies showed that troponin is associated with events more strongly over a shorter time horizon;[12] thus, temporal changes in the activity of pathways may also be relevant.

Furthermore, our results highlight that the pathophysiology of coronary disease and ischemic events is complex, involving the interplay of multiple pathophysiological pathways, and this complexity will likely not be adequately represented except by multiple biomarkers or potentially by biomarkers that reflect the intersection of more than one pathway. We found several biomarkers that were strongly associated with clinical outcomes that seemed to reflect the interplay of more than one pathway. For example, soluble CD40L plays a role in leukocyte-platelet interaction in acute coronary syndrome (ACS) and is a marker of platelet activation and

endothelial dysfunction; thus, it sits at the intersection of several pathways (endothelial function, thrombosis, inflammation, and atherosclerosis) in the pathogenesis of coronary disease and ACS. In our analyses, sCD40L was selected in all 3 fitted models for death/MI and death, with a stronger association in the death/MI model. This finding may reflect differences in the role of sCD40L or its associated pathways in these events and could be important to consider in further development of this biomarker.

Additionally, the MMPs are implicated in plaque instability and adverse ventricular remodeling post-MI and are associated with higher rates of death and heart failure.[13,14] MMP-3 was consistently associated with death/MI and death in all 3 fitted models. This is of particular interest as only 35% of our cohort had a history of MI or heart failure and only 15% presented with an acute MI. Additionally, the median ejection fraction of our cohort was 55%, and only 26.5% of patients had an NYHA functional class of ≥II. Therefore, plasma concentrations of MMP-3 may reflect the intersection of multiple pathways that underpin these clinical events and could be a candidate for development as a biomarker of risk for developing plaque rupture and left ventricular dysfunction prior to exhibiting any clinical signs or symptoms.

Metabolism

Without considering multiple biomarkers from multiple pathways simultaneously, including proteins from pathways that historically have not been as highly considered or published for their associations with cardiac outcomes, potentially important insights into pathophysiology and candidate biomarkers could be missed. For example, one of our interests has been in the role of regulatory hormones, such as insulin-like growth factors and binding proteins in the pathophysiology and outcomes of coronary disease, but IGFBP2, which was highly and independently associated with mortality in our study has received little previous attention, thus

raising the opportunity for novel biomarker development or new mechanistic understanding from additional study of IGFBP2 as a component of a multimarker panel.

In summary, we evaluated biomarkers from pathways that have been implicated in the development of CVD and that previously were associated with cardiovascular events; however, when considered in the context of other proteins (within the same pathway and from other pathways) and clinical factors, few remained significantly associated with outcomes. Of the 53 proteins studied, biomarkers from pathways of inflammation and atherosclerosis (IL-6, VCAM-1, and ICAM-1), extracellular matrix remodeling and endothelial dysfunction (MMP-3 and sCD40L), hemodynamic stress (NT-proBNP), and metabolism (IGFBP2) were consistently associated with death and death/MI.

Strengths and Limitations

Our results reflect associations; causality cannot be inferred and the ability of identified biomarker clusters to predict future events cannot be determined from our case-control design and must be confirmed in prospective studies. As suggested in a prior publication by Hlatky,[15] our approach is an early step in the biomarker-development process. However, these analyses provide insight into the potential utility of high throughput analytical techniques and statistical methods that simultaneously assess many potential protein biomarkers in the context of clinical features as a means to derive smaller panels of highly relevant proteins for focus in development of useful adjuncts to clinical risk assessment and stratified cardiovascular care.

We used the elastic net to select high-priority variables from a set of putative biomarkers of clinical events. The advantage of this strategy is the ability to develop a larger list of candidate variables for future analysis. To internally validate our findings and fully account for biomarkers with nonlinear associations, we also used the group LASSO technique to analyze our data as a

sensitivity analysis. The group LASSO selected a smaller number of biomarkers; however, most of the markers that were strongly associated with outcomes using the elastic net were also selected by the group LASSO.

Our patient population was a heterogeneous group presenting to a single center for suspected cardiovascular event or referred for elective coronary angiography; thus, the prevalence of cardiovascular risk factors and known coronary disease and our observations may not be representative of all populations with suspected coronary disease or more general populations. We were only able to match on age within 10 years. Because biomarker levels may vary with age, this may have influenced the associations (or lack thereof) that we observed, despite adjusting for age in models 2 and 3. Furthermore, we were not able to distinguish cardiovascular death from other causes of death in our database. While this does not invalidate our observed associations, we acknowledge that biomarker associations with mortality will likely vary by cause of death. Also, blood samples were only available at the time of coronary angiography, providing a "snapshot" protein profile at one point in time. The value of serial sampling for changes in biomarkers will need to be addressed in future work. Because medication data were not consistently available in this dataset, we could not adjust for differences in medication usage between cases and controls. Finally, we used available high-throughput analytical platforms to assess levels of multiple proteins simultaneously. Because of variability in assay performance characteristics, our results on an individual biomarker basis may have varied from those that were the same assay used as in literature-based studies. All of these limitations will need to be addressed in future prospective studies. However, we do not believe this overshadows the importance of the concept of considering multiple biomarkers simultaneously with appropriate statistical techniques to refine identification of the most relevant

biomarkers to examine in future studies.

Conclusions

High throughput analytical platforms and statistical methods allowing for simultaneous testing of multiple candidate biomarkers and clinical variables are useful in distilling large numbers of putative biomarkers to smaller panels for focus in the development of useful clinical tools. Future work in identifying novel biomarkers of risk should incorporate similar methods to allow for refined panels of proteins that are additive to readily available clinical prediction tools.

Acknowledgments: We would like to acknowledge the contributions of Terry Walker, PhD, and Peter Leitner, MS, for their assistance in the design and conduct of this project, and Morgan deBlecourt for her editorial assistance, which was provided as part of her regular duties as an employee of the Duke Clinical Research Institute.

Funding Sources: This study was supported by funding from the Measurement to Understand the Reclassification of Disease of Cabarrus and Kannapolis (MURDOCK) Study through a gift to Duke University from the David H. Murdock Foundation for Business and Culture; provision of research materials by diaDexus, Inc. (South San Francisco, CA); and grant number UL1RR024128-01 from the National Center for Research Resources, a component of the National Institutes of Health (NIH), and NIH Roadmap for Medical Research. None of the sponsors had a role in study design, data or sample collection, statistical analyses, interpretation of the data, writing of the study manuscript, or the decision to publish. CATHGEN, a biorepository from which the samples were obtained, is a self-funded sample and data repository that resides with the Duke Division of Cardiology, Department of Medicine.

Conflict of Interest Disclosures: Dr. Califf receives research grant funding from Amylin, BMS, Eli Lilly & Company, Janssen Research & Development LLC, Merck, Novartis, Schering-Plough, and Scios; consulting or other services (including CME) from Genentech, GSK, Janssen Research & Development LLC, Medscape LLC/heart.org, and Roche; and has equity in N30 Pharma and Portola. Dr. Newby receives research grant funding from the MURDOCK Study and received in-kind services from DiaDexus, Inc. in direct support of this study. Complete listings of Dr. Califf's, Dr. Granger's, and Dr. Newby's relationships with industry are available at https://www.dcri.org/about-us/conflict-of-interest. The other authors have no disclosures.

References:

1. Shah SH, Granger CB, Hauser ER, Kraus WE, Sun JL, Pieper K, et al. Reclassification of cardiovascular risk using integrated clinical and molecular biosignatures: Design of and rationale for the Measurement to Understand the Reclassification of Disease of Cabarrus and Kannapolis (MURDOCK) Horizon 1 Cardiovascular Disease Study. *Am Heart J*. 2010;160:371–379.

2. Bergstrahl EJ, Kosanke JL. Computerized matching of controls. Technical Report Series No. 56, Department of Health Science Research, Mayo Clinic, Rochester, MN. 1995.

3. Rosenbaum PR. Optimal matching for observational studies. *J Am Stat Assoc.* 1989;84:1024–1032.

4. Zou H, Hastie T. Regularization and variable selection via the elastic net. *J R Stat Soc*. 2005;67:301–320.

5. Yuan M, Lin Y. Model selection and estimation in regression with grouped variables. *J R Stat Soc*. 2006;68:49–67.

6. Albert MA, Danielson E, Rifai N, Ridker PM. Effect of statin therapy on C-reactive protein levels: the pravastatin inflammation/CRP evaluation (PRINCE): a randomized trial and cohort study. *JAMA*. 2001;286:64–70.

7. Morrow DA, Rifai N, Antman EM, Weiner DL, McCabe CH, Cannon CP, Braunwald E. C-reactive protein is a potent predictor of mortality independently of and in combination with troponin T in acute coronary syndromes: A TIMI 11A substudy. Thrombolysis in Myocardial Infarction. *J Am Coll Cardiol*. 1998;31:1460–1465.

8. Toss H, Lindahl B, Siegbahn A, Wallentin L. Prognostic influence of increased fibrinogen and C-reactive protein levels in unstable coronary artery disease. FRISC Study Group. Fragmin during Instability in Coronary Artery Disease. *Circulation*. 1997;96:4204–4210.

9. Ridker PM, Danielson E, Fonseca FA, Genest J, Gotto AM, Jr., Kastelein JJ, et al. Rosuvastatin to prevent vascular events in men and women with elevated C-reactive protein. *N Engl J Med*. 2008;359:2195–2207.

10. de Lemos JA, Drazner MH, Omland T, Ayers CR, Khera A, Rohatgi A, et al. Association of troponin T detected with a highly sensitive assay and cardiac structure and mortality risk in the general population. *JAMA*. 2010;304:2503–2512.

11. deFilippi CR, de Lemos JA, Christenson RH, Gottdiener JS, Kop WJ, Zhan M, Seliger SL. Association of serial measures of cardiac troponin T using a sensitive assay with incident heart failure and cardiovascular mortality in older adults. *JAMA*. 2010;304:2494–2502.

12. Westerhout CM, Fu Y, Lauer MS, James S, Armstrong PW, Al-Hattab E, et al. Short- and long-term risk stratification in acute coronary syndromes: the added value of quantitative ST-segment depression and multiple biomarkers. *J Am Coll Cardiol*. 2006;48:939–947.

13. Inoue T, Kato T, Takayanagi K, Uchida T, Yaguchi I, Kamishirado H, et al. Circulating matrix metalloproteinase-1 and -3 in patients with an acute coronary syndrome. *Am J Cardiol*. 2003;92:1461–1464.

14. Kelly D, Khan S, Cockerill G, Ng LL, Thompson M, Samani NJ, Squire IB. Circulating stromelysin-1 (MMP-3): A novel predictor of LV dysfunction, remodelling and all-cause mortality after acute myocardial infarction. *Eur J Heart Fail*. 2008;10:133–139.

15. Hlatky MA, Greenland P, Arnett DK, Ballantyne CM, Criqui MH, Elkind MS, et al. Criteria for evaluation of novel markers of cardiovascular risk:a scientific statement from the american heart association. *Circulation*. 2009;119:2408–2416.

DOI: 10.1161/CIRCGENETICS.113.000490

Table 1: Baseline Characteristics (Case vs. Control)

Baseline Characteristic	Case* (N=273)	Control (N=273)	P-value
Demographics			
Age	67.0 (58.0, 76.0)	62.0 (55.0, 70.0)	<0.001
Female	88 (32.2)	88 (32.2)	1.00
White	200 (73.3)	201 (73.6)	0.923
Clinical characteristics			
Height (cm)	173.0 (165.0, 180.0)	173.0 (168.0, 180.0)	0.040
Weight (kg)	80.0 (70.0, 92.0)	89.0 (75.0, 104.0)	<0.001
Body mass index (kg/m^2)	27.5 (24.2, 30.4)	28.7 (25.3, 33.1)	0.001
Heart rate (bpm)	73.0 (62.0, 84.0)	69.0 (60.0, 77.0)	0.004
Systolic blood pressure (mm Hg)	144.0 (125.0, 163.0)	147.0 (134.0, 162.0)	0.073
Diastolic blood pressure (mm Hg)	81.0 (70.0, 89.0)	84.0 (74.0, 94.0)	0.002
ECG characteristics			
Rhythm			
Normal sinus rhythm	199 (72.9)	240 (87.9)	<0.001
Atrial fibrillation/flutter	25 (9.2)	8 (2.9)	0.002
Other rhythm	30 (11.0)	16 (5.9)	0.031
Left bundle branch block	5 (1.8)	4 (1.5)	1.00
Left ventricular hypertrophy	53 (19.4)	42 (15.4)	0.214
QRS duration (ms)	95.0 (86.0, 117.0)	92.0 (84.0, 101.0)	0.001
QT interval (ms)	435.0 (412.0, 460.0)	422.0 (401.0, 446.0)	<0.001
ST-segment elevation	17 (6.2)	12 (4.4)	0.340
ST-segment depression	26 (9.5)	13 (4.8)	0.031
T-wave inversion	47 (17.2)	40 (14.7)	0.413
Non-specific ST-T wave changes	78 (28.6)	65 (23.8)	0.206
Q waves	68 (24.9)	46 (16.8)	0.021
Medical history			
Diabetes	108 (39.6)	77 (28.2)	0.005
Hypertension	192 (70.3)	193 (70.7)	0.925
Dyslipidemia	163 (59.7)	165 (60.4)	0.861
Smoking	148 (54.2)	120 (44.0)	0.017
Family history of coronary disease	112 (41.0)	73 (26.7)	<0.001
Prior MI	124 (45.4)	68 (24.9)	<0.001
Prior PCI	83 (30.4)	67 (24.5)	0.125
Prior CABG	88 (32.2)	59 (21.6)	0.005
Prior CVD	33 (12.1)	27 (9.9)	0.412
Prior PVD	46 (16.8)	15 (5.5)	<0.001
Carotid bruits	22 (8.1)	8 (2.9)	0.009
Valvular disease	16 (5.9)	7 (2.6)	0.055
Angina frequency (episode/week)	2.0 (0.0, 5.0)	3.0 (0.0, 5.0)	0.639
Angina during sleep	23 (8.4)	17 (6.2)	0.324
History of heart failure	119 (43.8)	63 (23.5)	<0.001
Charlson index	0.0 (0.0, 1.0)	0.0 (0.0, 0.0)	<0.001

DOI: 10.1161/CIRCGENETICS.113.000490

Baseline Characteristic	Case* (N=273)	Control (N=273)	P-value
NYHA class			<0.001
0	163 (59.7)	223 (81.7)	
I	10 (3.7)	5 (1.8)	
II	42 (15.4)	27 (9.9)	
III	47 (17.2)	14 (5.1)	
IV	11 (4.0)	4 (1.5)	
Renal disease	17 (6.2)	13 (4.8)	0.453
Clinical presentation			<0.001
Acute MI	57 (20.9)	26 (9.5)	
Outpatient	122 (44.7)	180 (65.9)	
Other	94 (34.4)	67 (24.5)	
Laboratories			
BUN (mg/dL)	20.0 (14.0, 28.0)	18.0 (14.0, 23.0)	0.002
Creatinine (mg/dL)	1.2 (0.9, 1.4)	1.0 (0.9, 1.2)	<0.001
Estimated CrCl (Cockroft-Gault) (mL/min)	67.7 (48.1, 94.4)	87.0 (68.4, 112.2)	<0.001
Hemoglobin (mg/dL)	12.8 (11.2, 14.0)	13.7 (12.5, 14.8)	<0.001
RDW (%)	14.4 (13.4, 15.5)	13.4 (12.9, 14.1)	<0.001
WBC (# cells/mL)	7.4 (6.0, 9.2)	7.0 (5.7, 8.4)	0.009
Sodium (mEg/L)	139.0 (137.0, 141.0)	140.0 (139.0, 142.0)	<0.001
Potassium (mEg/L)	4.2 (3.9, 4.5)	4.1 (3.9, 4.5)	0.611
Glucose (mg/dL)	106.0 (92.0, 142.0)	107.0 (92.0, 128.5)	0.915
Angiographic characteristics			
Duke index	45.0 (31.0, 77.0)	31.0 (20.0, 52.0)	<0.001
Number of diseased vessels			<0.001
0	62 (22.7)	88 (32.1)	
1	48 (17.6)	74 (27.1)	
2	52 (19.0)	45 (16.5)	
3	93 (34.1)	59 (21.6)	
LV ejection fraction	52.8 (38.0, 60.0)	57.6 (52.2, 65.9)	<0.001
Mitral regurgitation	25 (13.8)	14 (6.7)	0.020

*Case refers to patient with a death/MI event.
Data are presented as median (Q1, Q3) for continuous variables and as N (%) for categorical variables.
BUN - blood urea nitrogen; CABG - coronary artery bypass graft; CrCl - creatinine clearance; CVD - cardiovascular disease; ECG - electrocardiogram; LV - left ventricular; MI - myocardial infarction; NYHA - New York Heart Association; PCI - percutaneous coronary intervention; PVD - peripheral vascular disease; RDW - red cell distribution width; WBC - white blood cell

DOI: 10.1161/CIRCGENETICS.113.000490

Table 2: Biomarker Concentrations (Cases vs. Controls)

Protein Biomarker	Case* (N=273)	Control (N=273)	P-Value
SAA (ng/mL)	8581.6 (3395.4, 30,009.0)	4155.0 (2075.4, 9336.2)	<0.001
CRP (ng/mL)	6070.6 (2242.8, 18,786.8)	2521.8 (968.8, 6840.6)	<0.001
VCAM-1 (ng/mL)	471.2 (384.6, 590.4)	400.2 (335.2, 469.4)	<0.001
ICAM-1 (ng/mL)	270.0 (221.4, 339.6)	236.2 (199.8, 281.2)	<0.001
Thrombomodulin (ng/mL)	2.9 (2.4, 3.9)	2.7 (2.2, 3.3)	0.001
ICAM-3 (ng/mL)	2.8 (2.1, 3.6)	2.5 (2.0, 3.1)	0.002
E-Selectin (ng/mL)	16.7 (12.6, 24.6)	16.3 (12.5, 22.4)	0.421
P-Selectin (ng/mL)	83.2 (57.8, 119.3)	75.1 (48.7, 110.9)	0.019
LBP (ng/mL)	7413.4 (5234.8, 11,404.6)	5898.6 (4581.6, 8457.2)	<0.001
MMP-1 (ng/mL)	20.4 (10.4, 36.7)	17.4 (9.3, 28.4)	0.016
MMP-3 (ng/mL)	14.3 (8.9, 24.7)	9.6 (6.5, 14.5)	<0.001
MMP-9 (ng/mL)	98.0 (62.3, 150.2)	83.3 (58.1, 115.1)	0.003
TIMP-1 (ng/mL)	248.7 (190.0, 357.0)	200.7 (159.2, 261.8)	<0.001
bFGF (pg/mL)	47.8 (25.6, 71.9)	36.4 (19.6, 65.7)	0.002
sFlt-1 (pg/mL)	1371.2 (558.4, 4288.9)	1128.0 (413.8, 4956.9)	0.598
PlGF (pg/mL)	24.3 (18.0, 35.0)	23.2 (17.2, 31.8)	0.076
VEGF (pg/mL)	276.5 (182.3, 450.0)	278.1 (159.8, 436.8)	0.429
CK-MB (ng/mL)	3.8 (2.5, 6.1)	3.3 (2.3, 5.5)	0.025
Myoglobin (ng/mL)	17.5 (13.8, 21.8)	15.9 (12.7, 18.7)	0.001
TnI (ng/mL)	0.1 (0.0, 0.3)	0.1 (0.0, 0.2)	0.094
MCP-1 (pg/mL)	381.1 (303.4, 482.8)	330.9 (270.5, 433.1)	0.001
TNFα (pg/mL)	2.6 (1.9, 3.8)	2.1 (1.6, 2.7)	<0.001
IL-1β (pg/mL)	0.5 (0.4, 0.8)	0.5 (0.3, 0.8)	0.194
IL-2 (pg/mL)	1.3 (0.5, 2.2)	1.2 (0.4, 2.0)	0.146
IL-4 (pg/mL)	1.4 (0.6, 2.7)	1.6 (0.7, 2.6)	0.518
IL-6 (pg/mL)	5.8 (3.2, 12.0)	3.2 (1.8, 6.4)	<0.001
IL-10 (pg/mL)	5.2 (3.5, 8.7)	4.7 (3.1, 7.5)	0.044
M-CSF (pg/mL)	17.9 (11.6, 33.0)	12.0 (8.6, 18.0)	<0.001
G-CSF (pg/mL)	9.9 (6.9, 14.7)	10.1 (7.2, 14.2)	0.875
IL-1α (pg/mL)	0.0 (0.0, 0.0)	0.0 (0.0, 0.0)	0.657
IL-1ra (pg/mL)	42.7 (20.1, 77.3)	50.3 (21.6, 84.5)	0.139
IL-18 (pg/mL)	220.2 (173.2, 285.2)	198.2 (156.2, 258.5)	0.006
OPGN (pg/mL)	525.2 (395.3, 756.7)	420.9 (320.4, 584.9)	<0.001

DOI: 10.1161/CIRCGENETICS.113.000490

Protein Biomarker	Case* (N=273)	Control (N=273)	P-Value
Apolipoprotein A1 (μg/mL)	528.4 (449.5, 592.4)	525.2 (463.0, 608.8)	0.346
Apolipoprotein B (μg/mL)	20.2 (14.4, 29.1)	20.0 (14.0, 27.7)	0.742
Apolipoprotein E (μg/mL)	37.6 (29.6, 48.4)	36.0 (27.8, 47.9)	0.326
PDGF AA (ng/mL)	22.3 (13.0, 36.4)	22.4 (12.8, 35.4)	0.911
PDGF AB/BB (ng/mL)	71.5 (35.6, 110.2)	67.7 (37.0, 103.2)	0.655
RANTES (ng/mL)	127.6 (67.2, 196.0)	122.9 (74.5, 175.1)	0.697
MPO (ng/mL)	66.0 (30.5, 115.0)	64.5 (21.9, 123.0)	0.583
Total PAI-1 (ng/mL)	53.0 (37.5, 75.5)	47.5 (35.0, 68.0)	0.036
sCD40L (pg/mL)	6414.4 (3387.8, 34,833.7)	5541.5 (2695.8, 17,262.2)	0.019
TNFβ (pg/mL)	8.8 (4.2, 17.0)	8.0 (4.3, 13.5)	0.445
NT-proBNP (pg/mL)	417.0 (117.9, 1257.3)	85.2 (15.4, 318.3)	<0.001
GH (pg/mL)	102.5 (35.0, 427.0)	50.0 (20.6, 179.0)	<0.001
Fibrinogen (μg/mL)	3320.6 (2698.8, 4001.5)	2960.5 (2450.2, 3597.6)	<0.001
PAPP A (μg/mL)	0.0 (0.0,0.0)	0.0 (0.0, 0.0)	0.524
vWF (μg/mL)	14.5 (7.6,22.6)	10.0 (4.9, 17.5)	<0.001
tPA (pg/mL)	4035.8 (23,71.3, 7314.1)	3488.8 (2119.8, 6456.4)	0.180
D-Dimer (ng/mL)	579.9 (355.1, 896.8)	371.6 (252.4, 612.4)	<0.001
GDF-15 (pg/mL)	1545.4 (1052.6, 2909.2)	1017.0 (735.2, 1523.1)	<0.001
Lp-PLA2 PLAC® (ng/mL)	210.9 (176.3, 261.4)	206.3 (179.5, 237.0)	0.220
Lp-PLA2 CAM (nmol/mL/min)	126.6 (101.8, 153.2)	124.1 (105.1, 143.9)	0.476
IGFBP2 (ng/mL)	1095.1 (586.7, 1787.5)	532.5 (364.7, 1058.2)	<0.001

*Case refers to patient with a death/MI event.
Data are presented as median (Q1, Q3).
bFGF - basic fibroblast growth factor; CAM - colorimetric activity method; CK-MB - creatine-kinase MB; CRP - C-reactive protein; G-CSF - granulocyte colony-stimulating factor; GDF - growth differentiation factor; GH - growth hormone; ICAM - intercellular adhesion molecule; IGFBP2 - insulin-like growth factor binding protein-2; IL - interleukin; LBP - lipopolysaccharide-binding protein; Lp-PLA2 - lipoprotein-associated phospholipase A2; MCP - monocyte chemoattractant protein; MCSF - macrophage colony stimulating factor; MMP - matrix metalloproteinase; MPO - myeloperoxidase; NT-proBNP - N-terminal pro-B-type natriuretic peptide; OPGN - osteoprotegerin; PAI - plasminogen activator inhibitor; PAPP - pregnancy-associated plasma protein; PDGF - platelet-derived growth factor; PIGF - placental growth factor; RANTES - regulated on activation normal T cell expressed and secreted; SAA - serum amyloid A; sCD40L - soluble CD40 ligand; sFlt-1 - soluble Fms-like tyrosine kinase-1; TIMP - tissue inhibitor of metalloproteinase; TNF-α - tumor necrosis factor alpha; TnI - troponin I; tPA - tissue plasminogen activator; VCAM-1 - vascular cell adhesion molecule-1; VEGF - vascular endothelial growth factors; vWF - von Willebrand factor

DOI: 10.1161/CIRCGENETICS.113.000490

Table 3: Biomarkers Associated with Death/MI in Elastic Net Models 1 and 2

	Model 1		Model 2*	
	% of 500 bootstrapped samples selected	OR†	% of 500 bootstrapped samples selected	OR†
NT-proBNP	99.8	1.21	–‡	–
IGFBP2	99.8	1.30	82.2	1.25
IL-6	94.2	1.22	78.6	1.21
sCD40L	91.6	1.16	90.0	1.15
MMP-3	88.4	1.23	74.7	1.23
ICAM-1	87.4	1.83	80.2	1.71

*Model 2 was constrained to retain the following clinical variables predictive of death/MI: age, sex, weight, blood pressure, heart rate, smoking history, diabetes, presence of chest pain at presentation, New York Heart Association class, ejection fraction, atrial fibrillation, left bundle branch block, corrected QT interval, red cell distribution width, serum blood urea nitrogen, creatinine, hemoglobin, white blood cell count, Duke coronary artery disease index, and Charlson comorbidity index.
†Odds ratios calculated after variable selection to provide insight into direction of effect, not magnitude.
‡Indicates protein selected in <70% of bootstrapped samples for the respective model.
ICAM-1 - intercellular adhesion molecule-1; IGFBP2 - insulin-like growth factor binding protein-2; IL-6 - interleukin-6; MI - myocardial infarction; MMP-3 - matrix metalloproteinase-3; NT-proBNP - N-terminal pro-B-type natriuretic peptide; OR - odds ratio; sCD40L - soluble CD40 ligand

DOI: 10.1161/CIRCGENETICS.113.000490

Table 4: Biomarkers Associated with Death in Elastic Net Models 1 and 2

	Model 1		Model 2*	
	% of 500 bootstrapped samples selected	OR†	% of 500 bootstrapped samples selected	OR†
NT-proBNP	100	1.20	–‡	–
MMP-3	91.6	1.21	80.4	1.38
IL-6	90.4	1.16	79.8	1.30
IGFBP2	90.0	1.16	–	–
D-Dimer	85.0	1.22	–	–
VCAM-1	82.0	1.33	–	–
GDF-15	78.4	1.07	–	–
ICAM-1	77.2	1.68	–	–
sCD40L	76.4	1.12	74.9	1.10

*Model 2 was constrained to retain the following clinical variables predictive of death: age, sex, weight, blood pressure, heart rate, smoking history, diabetes, presence of chest pain at presentation, New York Heart Association class, ejection fraction, atrial fibrillation, left bundle branch block, left ventricular hypertrophy, corrected QT interval, red cell distribution width, serum sodium, blood urea nitrogen, creatinine, hemoglobin, white blood cell count, Duke coronary artery disease index, and Charlson comorbidity index.
†Odds ratios calculated after variable selection to provide insight into direction of effect, not magnitude.
‡Indicates protein selected in <70% of bootstrapped samples for the respective model.
GDF-15 - growth differentiation factor-15; ICAM-1 - intercellular adhesion molecule-1; IGFBP2 - insulin-like growth factor binding protein-2; IL-6 - interleukin-6; MI, myocardial infarction; MMP-3 - matrix metalloproteinase-3; NT-proBNP - N-terminal pro-B-type natriuretic peptide; OR - odds ratio; sCD40L - soluble CD40 ligand; VCAM-1 - vascular cell adhesion molecule-1

DOI: 10.1161/CIRCGENETICS.113.000490

Table 5: Biomarkers and Clinical Variables Associated with Death/MI and Death in Model 3

	Death/MI		Death	
	% of 500 bootstrapped samples	OR*	% of 500 bootstrapped samples	OR*
Candidate biomarkers				
VCAM-1	–†	–	73.5	1.11
MMP-3	89.2	1.19	89.0	1.21
IL-6	93.2	1.19	91.6	1.18
IGFBP2	97.6	1.17	76.2	1.12
sCD40L	97.4	1.20	80.2	1.13
NT-proBNP	91.0	1.10	93.0	1.09
ICAM-1	88.4	1.79	70.3	1.80
PlGF	76.0	0.94	70.7	0.95
sFlt-1	77.4	1.15	–	–
D-dimer	–	–	71.9	1.07
IL-1RA	70.5	0.90	–	–
Clinical predictors				
RDW	97.6	1.23	97.6	1.22
Age (per 5 yrs)	94.8	1.26	96.8	1.23
NYHA class	85.2	1.12	95.0	1.22
Diabetes	85.6	1.53	–	–
Creatinine (per 0.1 mg/dL)	70.3	1.06	86.2	1.11
Baseline Hgb (per 1 g/dL increase)	86.2	0.91	–	–
SBP (per 5 mmHg)	80.8	0.88	–	–
DBP (per 5 mmHg)	–	–	75.4	0.92
Ejection fraction (per 5% increase)	81.0	0.94	76.6	0.93
Smoking history	81.6	1.65	–	–
Weight (per 10 kg increase)	71.1	0.86	–	–
Duke index (per 10 units)	78.6	1.06	–	–
Sodium	NA	NA	70.1	0.95

*OR calculated after variable selection to provide insight into direction of effect, not magnitude.
†Indicates protein selected in <70% of bootstrapped samples for the respective model.
DBP - diastolic blood pressure; Hgb - hemoglobin; ICAM-1 - intercellular adhesion molecule-1; IGFBP2 - insulin-like growth factor binding protein-2; IL-1RA - IL-1 receptor antagonist; IL-6 - interleukin-6; MI - myocardial infarction; MMP-3 - matrix metalloproteinase-3; NT-proBNP - N-terminal pro-B-type natriuretic peptide; NYHA - New York Heart Association; OR - odds ratio; PIGF - placental growth factor; RDW - red cell distribution width; SBP - systolic blood pressure; sCD40L - soluble CD40 ligand; sFlt-1 - soluble Fms-like tyrosine kinase-1; VCAM-1 - vascular cell adhesion molecule-1

DOI: 10.1161/CIRCGENETICS.113.000490

Table 6: C-Indices with 95% Confidence Intervals for Clinical and Biomarker Models

Model	C-Index (95% CI)	Bias-Corrected C-Index	P-Value Comparing with M0
Death/MI			
M0: Clinical	0.795 (0.758–0.833)	0.759	–
M1: Proteins Only	0.781 (0.742–0.820)	0.775	0.83
M2: Proteins | Clinical	0.828 (0.793–0.863)	0.788	0.60
M3: Proteins + Clinical	0.824 (0.789–0.858)	0.795	0.65
Death			
M0: Clinical	0.800 (0.762–0.839)	0.754	–
M1: Proteins Only	0.785 (0.746–0.823)	0.767	0.83
M2: Proteins | Clinical	0.825 (0.789–0.861)	0.789	0.70
M3: Proteins + Clinical	0.816 (0.781–0.852)	0.795	0.73

CI - confidence interval; MI - myocardial infarction