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

Svati H. Shah, MD, MS, MHS,[a,b] Christopher B. Granger, MD,[a,c] Elizabeth R. Hauser, PhD,[b] William E. Kraus, MD,[a] Jie-Lena Sun, MS,[c] Karen Pieper, MS,[c] Charlotte L. Nelson, MS,[c] Elizabeth R. Delong, PhD,[c,d] Robert M. Califf, MD,[a,e] and L. Kristin Newby, MD, MHS[a,c] for the MURDOCK Horizon 1 Cardiovascular Disease Investigators[f] ***Durham, NC***

Background Clinical predictive models leave gaps in our ability to stratify cardiovascular risk. High-throughput molecular profiling promises to improve risk classification.

Methods Horizon 1 of the **M**easurement to **U**nderstand the **R**eclassification of **D**isease of **C**abarrus and **K**annapolis (MURDOCK) Study was conceived to apply emerging molecular techniques to existing data sets to characterize mechanistic diversity underlying complex human diseases, response to therapy, and prognosis. No previous studies have applied multiple, complementary molecular techniques in combination with well-developed clinical risk models to refine cardiovascular risk prediction. The MURDOCK Cardiovascular Disease Study will assess molecular profiles integrated with clinical data in "clinomic" profiles for cardiovascular risk classification.

Conclusion Herein, we describe the design of and rationale for the MURDOCK Cardiovascular Disease Study. (Am Heart J 2010;160:371-379.e2.)

Randomized clinical trials have identified several therapies that applied broadly to at-risk populations reduced the risk of myocardial infarction (MI) or death. Although such trials are integral to determining the safety and efficacy of drugs and devices, they define "average" responses over large cohorts of patients. "Personalized medicine," or perhaps better stated "stratified medicine," focuses on understanding the specific treatments and interventions most beneficial (and least harmful) in refined subgroups of patients. A more refined description of death or MI risk could facilitate tailored therapy and fuel discovery of new treatments. Blood-based biomarkers may serve this purpose. Novel molecular techniques (eg, genomics, proteomics, and metabolomics) facilitate high-throughput measurement of diverse biomarkers. Application and integration of biomarkers measured by these techniques may yield more refined risk assessment than traditional risk factors alone, helping to achieve the goal of stratified medicine and providing clearer understanding of underlying disease processes.

From the [a]Division of Cardiovascular Medicine, Duke University Medical Center, Durham, NC, [b]Duke Center for Human Genetics, Duke University Medical Center, Durham, NC, [c]Duke Clinical Research Institute, Duke University Medical Center, Durham, NC, [d]Department of Biostatistics and Bioinformatics, Duke University Medical Center, Durham, NC, and [e]Duke Translational Medicine Institute, Duke University Medical Center, Durham, NC.

[f]See online Appendix for a complete listing of MURDOCK Horizon 1 Cardiovascular Disease Investigators.

Submitted February 7, 2010; accepted June 24, 2010.

Reprint requests: L. Kristin Newby, MD, MHS, Duke Clinical Research Institute, PO Box 17969, Durham, NC 27715-7969.

E-mail: newby001@mc.duke.edu

0002-8703/$ - see front matter

© 2010, Mosby, Inc. All rights reserved.

doi:10.1016/j.ahj.2010.06.051

Current models for risk prediction are insufficient

The ability to discriminate risk using clinical characteristics and routine clinical laboratory testing is limited. The discriminative ability of the Framingham Risk Score (FRS), the most widely used clinical model for quantification of coronary heart disease (CHD) risk in the general population, is modest (c-index 0.69 in men and 0.72 in women).[1] The addition of high-sensitivity C-reactive protein (CRP) to the FRS model yielded a c-index of 0.81.[2] In the GRACE cohort, the c-index of a post–acute coronary syndrome (ACS) risk model that included

markers of myonecrosis and serum creatinine was also 0.81.[3] Still, opportunity remains to more effectively classify disease and estimate risk for clinical events.

Molecular technologies and molecular epidemiology of CHD

High-throughput molecular ("omics") technologies enable simultaneous, rapid measurement of thousands of DNA variants (genomics), RNA tags (transcriptomics), proteins (proteomics), and metabolites (metabolomics). Molecular signatures generated by these platforms are refining our understanding of pathophysiology and clinical care of common diseases.[4] Integrated with conventional clinical risk models, they promise to bring stratified cardiovascular care to fruition.

For example, genomewide association studies identified variants on chromosome 9p21 associated with CHD or MI.[5] The associations were modest; but because the variants are common, they represent significant population attributable risk. However, the functional and clinical implications of these intergenic variants are unclear. Furthermore, the presence of genetic variants alone does not specify that disease will be manifest. Thus, an individual's genotype alone is unlikely to provide risk stratification for clinical CHD events sufficient to guide individualized treatment.

RNA expression patterns, reflecting active transcription of genetic information, may be more descriptive of disease state and activity. Microarray analysis simultaneously determines expression levels of thousands of genes. Using peripheral blood RNA profiling, we found that the simultaneous expression pattern of 160 genes correlated with angiographic severity of coronary artery disease (CAD) and mirrored expression changes in atherosclerotic vascular wall.[6] Gene expression profiles also differentiated patients with angiographic CAD from controls, and a summary expression score of 14 genes was proportional to the extent of CAD.[7,8] We hypothesize that peripheral blood gene expression will also inform on risk for cardiovascular events.

Posttranscriptional modification via alternative splicing and other processes yields substantial protein diversity: roughly 4 to 50 proteins per gene and 100,000 to 1,000,000 total proteins. Protein biomarkers are more proximate to disease development, progression, and clinical events and should be complementary to clinical and genomic approaches. Clinicians are already accustomed to protein biomarkers in cardiology practice (eg, troponin, high-sensitivity CRP, and B-type natriuretic peptide). Troponin is a prototype biomarker of stratified cardiovascular medicine. Troponin defines MI,[9] predicts death or (re)MI among ACS patients,[10,11] and identifies which patients derive greatest benefit from certain therapies.[12,13]

Simple scores of multiple protein biomarkers provide additive, complementary information for MI diagnosis and risk stratification among ACS patients.[14,15] The next step is to combine protein biomarkers into more sophisticated high-throughput panels using large-scale "candidate protein" approaches and unbiased open proteomic platforms for novel protein discovery. In a proteomics discovery project, we identified both known and novel proteins characterizing patients with angiographic CAD.[16] Importantly, several differentially expressed proteins corresponded to pathways identified by peripheral blood RNA expression profiling, suggesting complementarity of these approaches that should be useful for predicting clinical events.

Changes in small-molecule metabolite levels provide an integrated "read-out" of genomic, transcriptomic, and proteomic variation. Unbiased metabolomic profiling has uncovered differences in CHD presence and severity[17]; however, most analytes are unidentified, challenging clinical and mechanistic interpretation. Using targeted, quantitative metabolomic profiling, we found high heritabilities for metabolites in humans and identified 2 metabolite factors independently associated with CAD.[18,19] In post hoc analyses, a third factor was associated with incident CHD events.

Need for integrated molecular and clinical models for cardiovascular risk

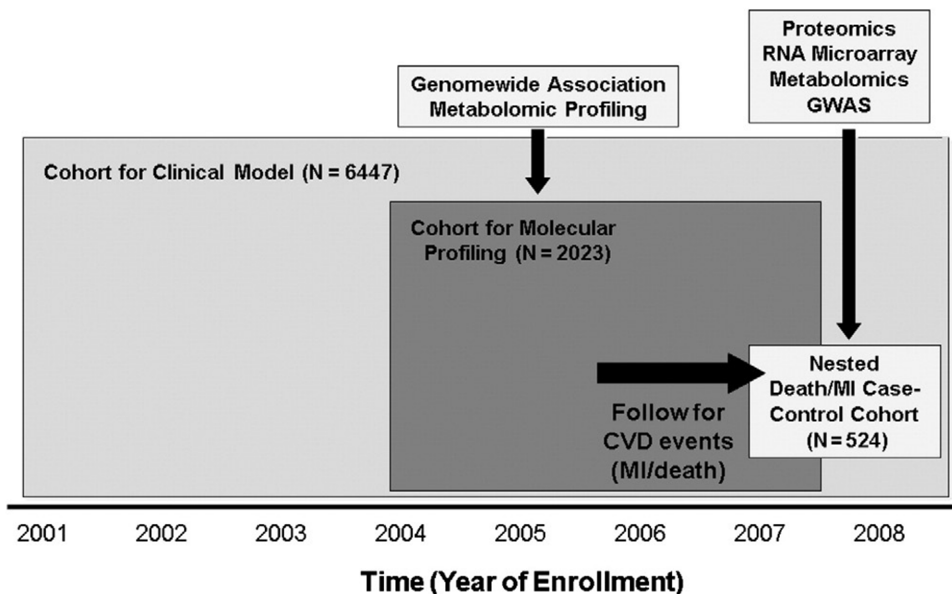
It is critical to develop general models predictive of "hard" cardiovascular end points (eg, MI and death) in which to assess the incremental contribution of molecular biosignatures. Most previous studies did not assess incremental utility of molecular testing in addition to risk factors that are easily measured during routine clinical assessment. We believe that the crucial next step in refining CHD event prediction is to integrate molecular biosignatures from multiple platforms with each other and, equally importantly, clinical predictors of CHD risk. On that backdrop, we designed the **M**easurement to **U**nderstand the **R**eclassification of **D**isease of **C**abarrus and **K**annapolis (MURDOCK) Horizon 1 Cardiovascular Disease Study to generate integrated clinical-molecular ("clinomic") profiles to better classify risk for future cardiovascular events.

Methods

The MURDOCK Study

The MURDOCK Study (http://www.dtmi.duke.edu/dtri/murdock), centered in Cabarrus County and Kannapolis, NC, is funded by a gift to Duke University from the David H. Murdock Foundation. Its primary objective is to rewrite the textbook of medicine by reclassifying health and disease using genomic technologies and electronic health records. Horizon 1, encompassing initial exploratory studies, uses existing

Figure 1



Overview of MURDOCK Cardiovascular Disease Study design.

biological specimen collections paired with well-annotated clinical data sets. The Horizon 1 Cardiovascular Disease Study will leverage biological samples collected in the CATHeterization GENetics (CATHGEN) biorepository at Duke University Medical Center (DUMC) (http://www.cathgen.duhs.duke.edu/index.htm) and clinical data from the Duke Databank for Cardiovascular Disease (DDCD).

The primary objectives of the MURDOCK Horizon 1 Cardiovascular Disease Study are to (1) identify molecular signatures that improve risk stratification for MI and death, (2) integrate molecular signatures from multiple platforms with readily available clinical information to form clinical-molecular ("clinomic") biosignatures of risk, and (3) explore interactions between treatments and baseline "omic" profiles in risk for subsequent clinical events.

In addition to the MURDOCK Study gift, the Horizon 1 Cardiovascular Study receives support from grant 1UL1 RR024128-01 from the National Center for Research Resources. The study was approved by the Institutional Review Board of DUMC.

The CATHGEN biorepository

The CATHGEN biorepository has collected biological samples on sequential consenting patients undergoing cardiac catheterization at DUMC since 2001 and is approved by the DUMC Institutional Review Board. Following informed consent, 50 mL of blood is collected from fasting patients through a femoral arterial sheath at the time of vascular access and immediately cooled to 4°C. Plasma is prepared from EDTA tubes, centrifuged at 4°C and 1,500*g* for 10 minutes within 30 minutes of collection, separated into 0.5-mL aliquots, and stored at −80°C. Two 2.5-mL PAXgene RNA tubes, processed according to manufacturer's recommendations (http://www.preanalytix.com), and three 7.5-mL EDTA tubes for DNA are stored at −80°C.

The DDCD

Clinical information comes primarily from the DDCD. Available data include (1) symptom histories; (2) clinical characteristics and medical history; (3) angiographic data; and (4) in most subjects, fasting chemistry and lipid profile data within 1 year preceding cardiac catheterization. Using medical record numbers or other specific identifiers, we can query other Duke databases through the Decision Support Repository for additional clinical and administrative data. We have now added a link to medication data at the time of sampling to the CATHGEN registry. Using the electrocardiogram database (TraceMaster) supplemented by manual review of electrocardiograms, we also added electrocardiographic parameters for the entire cohort studied in this project. Access to medical records is available to clarify details not accessible from electronic sources.

All patients in the DDCD with a coronary artery stenosis >50% have routine follow-up at 6 months and 1 year postprocedure, then yearly thereafter. Follow-up includes mortality (verified via National Death Index search and supplemented by Social Security Death Index search), MI, stroke, rehospitalization, coronary revascularization procedures, smoking, and medication use. We also recently completed follow-up of all patients in CATHGEN who did not meet DDCD criteria for routine postprocedural follow-up. For this study, longitudinal follow-up was 93% complete.

Inclusion and exclusion criteria

Figure 1 displays the Horizon 1 MURDOCK Cardiovascular Disease Study design. Clinical event models will be constructed on 6,447 CATHGEN subjects enrolled between January 1, 2001, and November 14, 2007, who met the inclusion/exclusion criteria shown in Table I and Figure 2. For generalizability,

Table I. Inclusion and exclusion criteria

Inclusion criteria	Exclusion criteria
• Age >18 y	• Primary pulmonary hypertension or severe lung disease
• Consent for CATHGEN biorepository	• Cardiac transplantation
• Primary reason for catheterization is concern for ischemic heart disease	• Other solid organ transplant
• Complete coronary angiogram data available	• Peripheral vascular intervention only
	• Right heart catheterization only
	• Advanced heart failure (NYHA class IV with systolic dysfunction [left ventricular ejection fraction <35%] at cardiac catheterization)
	• Congenital heart disease

NYHA, New York Heart Association.

exclusion criteria focused only on major illnesses that would overwhelm, obscure, or confound associations of molecular signatures with CHD events. The molecular profiling cohort comprises 2,023 sequential patients with all biological sample types available. We will perform targeted metabolomic profiling in all samples from this cohort and transcriptomic and targeted proteomic profiling on a nested case-control set of 500 subjects (250 cases with death or MI any time after enrollment cardiac catheterization and 250 sex-, race-, and age-matched controls without death or MI through >2 years of follow-up). Discovery proteomics will be performed on a further 30-case/30-control subset randomly selected from the case-control set used for targeted profiling. Through separate funding (National Institutes of Health–National Heart, Lung, and Blood Institute 1RC2HL101621-01), we will complete expression profiling and perform genomewide association studies on the entire 2,000-patient cohort. When resources are available, we will also complete targeted protein profiling, guided by results from initial targeted and open-platform subsets.

Table II presents baseline clinical characteristics of the overall and molecular profiling cohorts, which reflect a typical population referred for cardiac catheterization, including high burdens of diabetes, obesity, hypertension, smoking, and dyslipidemia. The characteristics of the clinical cohort and the molecular profiling subset were similar.

End points

The primary end points are all-cause mortality and death or MI. *Myocardial infarction* was defined as creatine kinase–MB or troponin I or T levels higher than the upper limit of normal in patients with chest pain, cardiac arrest, or other symptoms suggestive of cardiac ischemia. Overall, median follow-up was 4.2 (interquartile range 2.3-5.9) years, 1,284 patients died or had MI, and 1,037 died. In the molecular subset (n = 2,023), median follow-up was 3.1 (interquartile range 2.1-3.8) years, 294 died or had MI, and 234 died.

Clinical models for cardiovascular event prediction

Cox proportional hazards models for death and death or MI will be developed to identify baseline clinical factors predictive of long-term outcomes in patients with suspected CHD undergoing cardiac catheterization. These clinical models will serve as a foundation for evaluation of incremental biomarker contributions in assessing risk and a base for integration of molecular factors from various platforms into a common risk profile. Secondarily, the clinical risk models will update older DDCD models[20-23] to (1) reflect current treatment and outcome patterns; (2) include patients with minimal disease (ie, medically treated), no disease, and left main disease, who were excluded in previous models primarily used to assess treatment-related survival among revascularization candidates; and (3) adjust for and evaluate treatment using landmark models.

Predictor variables will be selected using stepwise and backward selection at an entry criterion of $P < .1$ and retained at $P < .05$. Candidate variables include demographics, disease-specific illness severity, electrocardiographic variables, and standard laboratory findings. Additional considerations are prior evidence of association with the outcome of interest, clinician insight, and prevalence of ≥1% in the data set. Linearity of relationships between continuous variables and clinical outcomes will be tested using restricted cubic splines. Multiple imputation methods will account for additional variability in the parameter estimates for clinical variables with missing data. Outlier values for continuous variables will be set to missing before the imputation process. Variables that violate the proportional hazards assumption will be eliminated or, if clinically important, used as stratification factors. Biologically plausible potential interactions among clinically prespecified subgroups will be considered. Models will be internally validated using bootstrapping techniques. Strength of association of each variable with the clinical end point will be assessed by its Wald χ^2 value. The discriminative ability of each clinical model will be assessed by the modified c-index for survival data; and their overall strength, by the model Wald χ^2. Calibration curves will be prepared to assess model calibration.

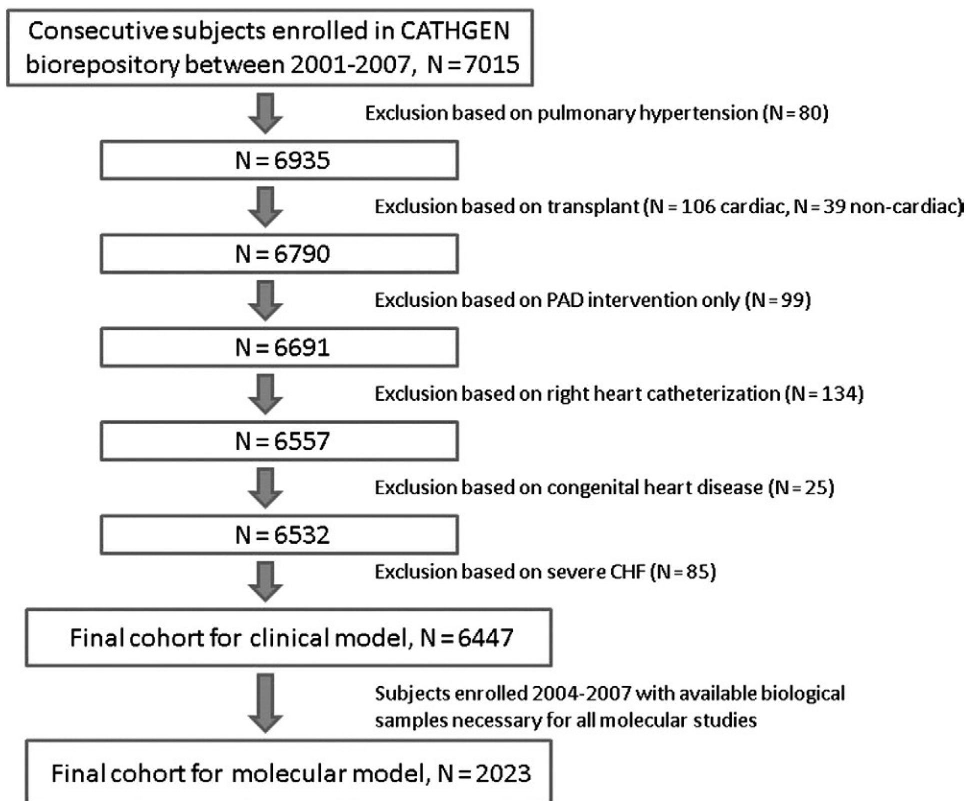
Cox proportional hazards models incorporating landmark modeling techniques will examine the contribution of treatment (medical, percutaneous coronary intervention, and coronary artery bypass surgery) to clinical outcomes.

RNA expression profiling

RNA will be extracted from stored frozen PAXGene tubes (PreAnalytiX GmbH; Becton, Dickinson and Company, Franklin Lakes, NJ) using previously used techniques for an initial globin reduction step and RNA extraction followed by a quality assessment step, providing a measure of RNA integrity of each sample. We will use Affymetrix U133 2.0 Plus arrays (Affymetrix, Inc., Santa Clara, CA) for RNA expression profiling. After running microarray chips, normalization will be performed using robust multiarray analysis, with background adjustment and quantile normalization to reduce nonbiological noise. We will then evaluate the coherence between technical replicates and assess for batch effects. If batch effects are present, we will apply empirical Bayesian batch normalization using the ComBat program, a well-accepted method for handling such effects.[24]

To address the large number of variables generated by RNA profiling, we will use factor analysis with imposition of a sparseness constraint to identify clusters of coexpressed genes in the case-control data set. Sparse probit-regression classifiers with full posteriors will be built to identify gene clusters associated with cardiovascular events (ie, best differentiating event cases from controls). We will also perform bootstrapping

Figure 2



Study population for MURDOCK Cardiovascular Disease Study.

with replacement to assess classification error and estimates of variable importance. In secondary analyses, we will use "biased" approaches for identification of gene expression tags (eg, hierarchical cluster analysis) to identify patterns best differentiating cases from controls. We will examine the genes and metagenes that are highly associated with clinical events to determine biological pathways involved in plaque instability and will adjust associations for clinical predictors of outcome.

Proteomics profiling

Targeted proteomic profiling assessing approximately 50 candidate proteins prioritized from the literature and our previous work (Appendix online, Supplement 2) will be performed on the same nested case-control subset used for RNA expression profiling. For this, we will use the high-throughput Meso Scale Discovery platform at the David H. Murdock Research Institute Core Laboratory facility in Kannapolis, NC. The Luminex platform will be used for protein assays not available through Meso Scale Discovery. Lp-PLA2 analysis will be performed at diaDexus laboratories (South San Francisco, CA). We will also perform open platform (discovery) proteomic profiling (all proteins present in a sample are resolved spectrally on a relative scale based on mass) in a 30-case/30-control subset using liquid chromatography/mass spectrometry techniques.

We will create panels of several proteins measured simultaneously (not simply a single protein in isolation) for use in risk discrimination. Univariate nonparametric testing will compare levels of each protein between case and control groups. Similar methods as described for RNA analyses will be used to identify coexpressed proteins and assess differences in these protein factors between cases and controls.

Metabolomic profiling

Targeted, quantitative metabolomic profiling using a mass spectrometry–based platform and standard analytical chemistry and immunoassay approaches will be performed using previously described protocols.[25] Metabolites (approximately 80) include lipids, 15 amino acids, hormones of satiety and hunger (peptide YY, agouti-related protein, ghrelin, and leptin), 45 acylcarnitine species (byproducts of mitochondrial free fatty acid, carbohydrate, and amino acid catabolism), ketones, nonesterified fatty acids, β-hydroxybutyrate, human growth hormone, adiponectin, insulin growth factor, and insulin-like growth factor binding proteins 1, 2, and 3.

Principal components analysis will reduce the large number of collinear metabolites into a smaller number of uncorrelated clusters. Univariable and multivariable associations of principal components analysis–derived metabolite clusters with clinical end points will be assessed in Cox proportional hazards models that include predictor variables identified in the overall cohort.

Table II. Baseline clinical characteristics

	Overall MURDOCK cardiovascular cohort (N = 6447)	Molecular profiling cohort (n = 2023)
Demographics		
Age, y	61.0 (52.0, 70.0)	62.0 (54.0, 70.0)
Race (white)	75.7%	73.4%
Sex (female)	37.0%	38.4%
Year of cardiac catheterization		
2001	10.3%	0%
2002	17.3%	0%
2003	20.4%	0%
2004	12.5%	14.6%
2005	16.1%	46.4%
2006	7.3%	22.3%
2007	16.2%	16.6%
Clinical characteristics		
Height, cm	173.0 (165.0, 180.0)	173.0 (165.0, 180.0)
Weight, kg	86.0 (74.0, 100.0)	86.0 (74.0, 101.0)
Heart rate	69.0 (60.0, 79.0)	70.0 (60.0, 79.0)
Systolic blood pressure, mm Hg	143.0 (129.0, 160.0)	143.0 (129.0, 161.0)
Diastolic blood pressure, mm Hg	80.0 (71.0, 89.0)	82.0 (73.0, 91.0)
Body mass index, kg/m^2	29.0 (25.5, 33.3)	29.0 (25.4, 33.3)
<18.5	1.3%	1.4%
18.5-25	20.8%	20.8%
>25-30	35.7%	35.3%
>30-35	24.0%	23.2%
>35-40	9.7%	10.3%
>40	8.5%	8.8%
Baseline ECG		
Rhythm		
NSR	77.6%	88.8%
Atrial fibrillation/flutter	4.3%	4.2%
Other	18.1%	11.0%
Left bundle-branch block	3.3%	2.3%
Left ventricular hypertrophy	15.1%	15.6%
QRS duration	90.0 (83.0, 101.0)	92.0 (83.0, 103.0)
QTc interval	420.0 (399.0, 444.0)	425.0 (404.0, 447.0)
ST-segment elevation	5.6%	5.2%
ST-segment depression	6.3%	7.9%
T-wave inversion	15.9%	16.7%
Nonspecific ST-T wave changes	22.0%	21.9%
Q waves	18.1%	17.7%
Medical history		
Diabetes mellitus	26.7%	28.9%
Hypertension	69.1%	67.5%
Dyslipidemia	60.8%	59.6%
Smoking	50.7%	47.2%
Family history of coronary disease	37.8%	36.1%
Prior myocardial infarction	29.4%	27.5%
Prior percutaneous coronary intervention	22.0%	22.1%
Prior coronary artery bypass grafting	19.4%	20.8%
Prior cerebrovascular disease	8.3%	8.6%
Peripheral vascular disease	7.5%	7.4%
Carotid bruits	5.5%	4.3%
Valvular disease	4.2%	5.1%
History of angina		
Angina frequency (episodes/wk)	2.0 (0.0, 5.0)	2.0 (0.0, 4.0)
Angina during sleep	9.1%	7.7%
History of heart failure	24.4%	26.2%
Noncardiac Charlson Index	0.0 (0.0, 0.0)	0.0 (0.0, 0.0)
Heart failure severity (NYHA class)		
0	80.4%	78.0%
1	1.7%	1.8%
2	7.4%	9.8%
3	8.5%	8.4%

Table II. (*continued*)

	Overall MURDOCK cardiovascular cohort (N = 6447)	Molecular profiling cohort (n = 2023)
4	2.2%	1.8%
Renal disease	1.8%	2.4%
Clinical presentation before enrollment catheterization		
Stable outpatient	54.4%	59.7%
Acute myocardial infarction	13.0%	10.6%
Inpatient, not myocardial infarction	32.6%	29.6%
Baseline laboratories		
Blood urea nitrogen	16.0 (13.0, 21.0)	17.0 (13.0, 21.0)
Serum creatinine	1.0 (0.9, 1.2)	1.0 (0.9, 1.2)
Estimated creatinine clearance (Cockroft-Gault)	88.9 (66.2, 114.7)	85.0 (62.9, 110.9)
Hemoglobin	13.6 (12.3, 14.7)	13.4 (12.2, 14.6)
Red cell distribution width	13.5 (12.9, 14.3)	13.6 (13.0, 14.4)
White blood cell count	7.2 (5.9, 8.8)	7.2 (5.9, 8.8)
Sodium	140.0 (138.0, 141.0)	140.0 (138.0, 142.0)
Potassium	4.1 (3.9, 4.4)	4.1 (3.9, 4.4)
Glucose	103.0 (92.0, 126.0)	106.0 (93.0, 130.0)
Baseline angiographic characteristics and left ventricular function		
Duke CAD severity index	31.0 (0.0, 58.0)	31.0 (0.0, 52.0)
No. of diseased vessels		
0	37.1%	38.6%
1	24.0%	23.7%
2	16.9%	16.3%
3	22.0%	21.4%
Left ventricular ejection fraction	57.5 (50.4, 65.0)	57.6 (51.0, 65.5)
Mitral regurgitation	14.2%	13.8%

Continuous variables are displayed as median (25th, 75th percentiles). Discrete variables are displayed as percentages of the column total N. *ECG*, Electrocardiogram; *NSR*, normal sinus rhythm.

Considerations of sample size, power, and population stratification

Explicit power calculations for the proposed samples sizes and analytic strategies are difficult. For example, for gene expression analyses, we cannot calculate power in the type of analysis we propose; but using a 1-way, 2-sided analysis of variance to detect changes in the 90% least variable genes ($\sigma \leq 1.1985$), with 250 cases and 250 controls, we are adequately powered to detect 1.5-fold changes at an α of .001. Similar effects are expected in both the metabolomic and proteomic data. Multiple comparisons will be addressed through multidimensional data reduction techniques.

Fundamental principles of predictive modeling support that for each variable included in a multivariable model, there should be 10 to 20 end points available for modeling. The “large p/little n” challenges generated by high-throughput genomic technologies are not immune to this general statistical rule. Assuming a clinical predictive model with 15 clinical variables and that a dozen or more multicomponent profiles will be generated for each platform, to have at least 10 end point events for each variable, at least 250 clinical events are required. These assumptions form the foundation for our nested case-control data set for protein and RNA expression analyses.

Recent studies used “omics” platforms with much smaller sample sizes. We will explicitly evaluate the effects of small samples sizes in the context of high-dimensional molecular data, analyzing the results for 100 randomly selected 30-case/30-control experiments, to assess the number of times these small case-control cohorts yield the same “factors” as each other and as in the overall main analysis. If there is consistency, these sensitivity analyses will be repeated with additional case-control sizes to determine if a “threshold effect” exists.

Race-specific differences in molecular profiles may exist (as shown for genetic variants). In exploratory analyses, we will evaluate for such race-specific variations and, if discovered, adjust our models accordingly. For any future genetic studies, we will stratify by race to address population stratification.

Assessing incremental predictive capability of molecular profiles

Evaluation of molecular biomarkers for use in clinical disease management cannot be dissociated from easily measured

clinical characteristics and routine laboratory values. Methods to assess the incremental contribution of new biomarkers are in continual development. Importantly, statistical significance of associations does not imply clinical significance. Furthermore, comparison of model strength with parameters such as changes in the log likelihoods between models with and without new biomarkers may give an incomplete picture of biomarker utility. In considering stratified medicine, which aims to predict at the small group to individual level, traditional receiver operating characteristics curve analysis (which gauges incremental risk at a population level) also may not be an appropriate metric. Increasingly, risk reclassification analysis is used to assess incremental predictive capability of molecular biomarkers.[26] Such analyses may provide a better metric of how biomarker information shifts an individual's estimated risk of events when added to traditional risk models.

We will assess the independent contribution of molecular risk factors in Cox proportional hazards models constrained to retain all clinical variables significant in the original clinical models. In each case, model calibration will be reassessed. We will examine multiple parameters including changes in model discrimination using the c-index and overall changes in global model strength as assessed by changes in log likelihoods between models with and without new factors. In addition, we will determine the degree to which molecular markers correctly reclassify a subject's risk category from that predicted by the clinical models. Our primary measure will be the net reclassification improvement (a measure of correctly up- and down-classifying risk relative to the base model and actual events). We will empirically determine risk thresholds by survey of multiple cardiovascular experts and assess the change in risk classification using these thresholds. In addition, we will examine reclassification relative to the FRS and GRACE Risk Score, although we acknowledge that neither cohort perfectly reflects the characteristics of our population.

Secondarily, we will integrate the diverse molecular platforms with the goal of identifying pathways mediating cardiovascular risk, using both "unbiased" (no a priori knowledge of biological pathways is used for identification of cross-platform signatures) and "biased" (publicly available databases are used to identify pathways associated with disease state by representation in multiple molecular platforms) methods.

The authors are solely responsible for the design and conduct of this study, all study analyses, the drafting and editing of this paper, and its final contents.

Discussion

Stratified cardiovascular care necessitates investigation of integrated clinical and molecular profiles of cardiovascular risk in carefully phenotyped populations. The MURDOCK Horizon 1 Cardiovascular Disease Study will integrate multiple molecular platforms with well-developed clinical models to identify novel "clinomic" signatures that we believe will provide more refined cardiovascular risk assessment. This will lead ultimately to new risk classification tools, tailored regimens for treatment and prevention of CHD, and improved clinical outcomes. These same techniques will also enhance our understanding of mechanisms of CHD and potentially lead to novel therapeutic targets that fuel development of new treatments.

References

1. Wilson PW, D'Agostino RB, Levy D, et al. Prediction of coronary heart disease using risk factor categories. Circulation 1998;97: 1837-47.
2. Ridker PM, Rifai N, Rose L, et al. Comparison of C-reactive protein and low-density lipoprotein cholesterol levels in the prediction of first cardiovascular events. N Engl J Med 2002;347:1557-65.
3. Eagle KA, Lim MJ, Dabbous OH, et al. A validated prediction model for all forms of acute coronary syndrome: estimating the risk of 6-month postdischarge death in an international registry. JAMA 2004;291:2727-33.
4. Potti A, Mukherjee S, Petersen R, et al. A genomic strategy to refine prognosis in early-stage non–small-cell lung cancer. N Engl J Med 2006;355:570-80.
5. Helgadottir A, Thorleifsson G, Manolescu A, et al. A common variant on chromosome 9p21 affects the risk of myocardial infarction. Science 2007;316:1491-3.
6. Sinnaeve PR, Donahue MP, Grass P, et al. Gene expression patterns in peripheral blood correlate with the extent of coronary artery disease. PLoS ONE 2009;e7037:4.
7. Wingrove JA, Daniels SE, Sehnert AJ, et al. Correlation of peripheral-blood gene expression with the extent of coronary artery stenosis. Circulation: Cardiovascular Genetics 2008;1:31-8.
8. Kraus WE, Rosenberg S, Elashoff MR, et al. Development and multi-center validation of a blood-based gene expression test for diagnosis of obstructive CAD in non-diabetic patients [abstract]. Circulation 2009;120:S563.
9. Thygesen K, Alpert JS, White HD, et al. Universal definition of myocardial infarction. Circulation 2007;116:2634-53.
10. Newby LK, Christenson RH, Ohman EM, et al. Value of serial troponin T measures for early and late risk stratification in patients with acute coronary syndromes. The GUSTO-IIa Investigators. Circulation 1998; 98:1853-9.
11. Ohman EM, Armstrong PW, Christenson RH, et al. Cardiac troponin T levels for risk stratification in acute myocardial ischemia. GUSTO IIA Investigators. N Engl J Med 1996;335:1333-41.
12. Morrow DA, Cannon CP, Rifai N, et al. Ability of minor elevations of troponins I and T to predict benefit from an early invasive strategy in patients with unstable angina and non–ST elevation myocardial infarction: results from a randomized trial. JAMA 2001;286: 2405-12.
13. Newby LK, Ohman EM, Christenson RH, et al. Benefit of glycoprotein IIb/IIIa inhibition in patients with acute coronary syndromes and troponin t-positive status: the paragon-B troponin T substudy. Circulation 2001;103:2891-6.
14. Newby LK, Storrow AB, Gibler WB, et al. Bedside multimarker testing for risk stratification in chest pain units: the chest pain evaluation by creatine kinase–MB, myoglobin, and troponin I (CHECKMATE) study. Circulation 2001;103:1832-7.
15. Sabatine MS, Morrow DA, de Lemos JA, et al. Multimarker approach to risk stratification in non-ST elevation acute coronary syndromes: simultaneous assessment of troponin I, C-reactive protein, and B-type natriuretic peptide. Circulation 2002;105:1760-3.
16. Donahue MP, Rose K, Hochstrasser D, et al. Discovery of proteins related to coronary artery disease using industrial-scale proteomics analysis of pooled plasma. Am Heart J 2006;152:478-85.

17. Brindle JT, Antti H, Holmes E, et al. Rapid and noninvasive diagnosis of the presence and severity of coronary heart disease using 1H-NMR-based metabonomics. Nat Med 2002;8:1439-44.
18. Shah SH, Hauser ER, Bain JR, et al. High heritability of metabolomic profiles in families burdened with premature cardiovascular disease. Mol Syst Biol 2009;5:258.
19. Shah SH, Bain JR, Muehlbauer MJ, et al. Association of a peripheral blood metabolic profile with coronary artery disease and risk of subsequent cardiovascular events. Circulation Cardiovasc Gen 2010;3:207-14.
20. Califf RM, Harrell Jr FE, Lee KL, et al. The evolution of medical and surgical therapy for coronary artery disease. A 15-year perspective. JAMA 1989;261:2077-86.
21. Jones RH, Kesler K, Phillips III HR, et al. Long-term survival benefits of coronary artery bypass grafting and percutaneous transluminal angioplasty in patients with coronary artery disease. J Thorac Cardiovasc Surg 1996;111:1013-25.
22. Mark DB, Nelson CL, Califf RM, et al. Continuing evolution of therapy for coronary artery disease. Initial results from the era of coronary angioplasty. Circulation 1994;89:2015-25.
23. Smith PK, Califf RM, Tuttle RH, et al. Selection of surgical or percutaneous coronary intervention provides differential longevity benefit. Ann Thorac Surg 2006;82:1420-8.
24. Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostatistics 2007;8:118-27.
25. Ferrara CT, Wang P, Neto EC, et al. Genetic networks of liver metabolism revealed by integration of metabolic and transcriptional profiling. PLoS Genet 2008;e1000034:4.
26. Pencina MJ, D'Agostino Sr RB, D'Agostino Jr RB, et al. Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. Stat Med 2008;27:157-72.

Appendix

Supplement 1: MURDOCK Horizon 1 Cardiovascular Disease Study investigators and affiliations

Robert M. Califf, MD, Duke Translational Medicine Institute, Duke University

Lawrence Carin, PhD, Department of Electrical and Computer Engineering, Duke University

Marie Davidian, PhD, Department of Statistics, North Carolina State University

Elizabeth R. Delong, PhD, Department of Biostatistics and Bioinformatics, Duke University

Christopher B. Granger, MD, Duke Clinical Research Institute, Duke University

Elizabeth R. Hauser, PhD, Center for Human Genetics, Duke University

William E. Kraus, MD, Department of Medicine, Duke University

Joseph Lucas, PhD, Institute for Genome Sciences and Policy, Duke University

Arthur Moseley, PhD, Institute for Genome Sciences and Policy, Duke University

Charlotte L. Nelson, MS, Duke Clinical Research Institute, Duke University

L. Kristin Newby, MD, MHS, Duke Clinical Research Institute, Duke University

Christopher B. Newgard, PhD, Director, Sarah W. Stedman Center for Nutrition and Metabolism, Duke University

Karen Pieper, MS, Duke Clinical Research Institute, Duke University

Svati H. Shah, MD, MS, MHS, Center for Human Genetics, Duke University

Jie-Lena Sun, MS, Duke Clinical Research Institute, Duke University

Butch Tsiatis, PhD, Department of Statistics, North Carolina State University

Terry Walker, PhD, Duke Translational Research Institute, Duke University

Supplement 2: Targeted proteins to be measured in the MURDOCK Cardiovascular Disease Study

Pathway	Protein	Platform
Inflammatory	Monocyte chemoattractant protein 1	Meso Scale Discovery
	Tumor necrosis factor α	Meso Scale Discovery
	Tumor necrosis factor β	MilliPlex
	Interleukin 1β	Meso Scale Discovery
	Interleukin 1α	Meso Scale Discovery
	Interleukin 1 RA	Meso Scale Discovery
	Interleukin 2	Meso Scale Discovery
	Interleukin 4	Meso Scale Discovery
	Interleukin 6	Meso Scale Discovery
	Interleukin 10	Meso Scale Discovery
	Interleukin 18	Meso Scale Discovery
	Macrophage colony-stimulating factor	Meso Scale Discovery
	Granulocyte colony-stimulating factor	Meso Scale Discovery
	Soluble intracellular adhesion molecule 1	Meso Scale Discovery
	Vascular intracellular adhesion molecule 1	Meso Scale Discovery
	Intercellular adhesion molecule 3	Meso Scale Discovery
	CRP	Meso Scale Discovery
	Serum amyloid A	Meso Scale Discovery
	P-selectin	Meso Scale Discovery
	E-selectin	Meso Scale Discovery
	Chemokine ligand 5 (aka *RANTES*)	MilliPlex
	Myeloperoxidase	MilliPlex
	CD 40 ligand	MilliPlex
	Lipoprotein-associated phospholipase A2	diaDexus
Thrombosis	Thrombomodulin	Meso Scale Discovery
	Plasminogen activator inhibitor 1 (active and total)	MilliPlex
	Fibrinogen	MilliPlex
	von Willebrand factor	American Diagnostics
	Tissue plasminogen activator	Assay Pro
	D-dimer	American Diagnostics
Vascular/endothelial	FMS-like tyrosine kinase 1	Meso Scale Discovery
	Vascular endothelial growth factor	Meso Scale Discovery
	Phosphatidylinositol-glycan biosynthesis class F protein	Meso Scale Discovery
	Basic fibroblast growth factor	Meso Scale Discovery
	Growth differentiation factor 15	R&D

Supplement 2 (*continued*)

Pathway	Protein	Platform
Myocardial proteins	Troponin I	Meso Scale Discovery
	Creatine kinase–MB	Meso Scale Discovery
	Myoglobin	Meso Scale Discovery
Extracellular matrix	Matrix metalloproteinase 1	Meso Scale Discovery
	Matrix metalloproteinase 3	Meso Scale Discovery
	Matrix metalloproteinase 9	Meso Scale Discovery
	Tissue inhibitor of metalloproteinase 1	Meso Scale Discovery
Lipid derived	Apolipoprotein A1	MilliPlex
	Apolipoprotein B	MilliPlex
	Apolipoprotein E	MilliPlex
Miscellaneous	Osteoprotegerin	Meso Scale Discovery
	Alkaline phosphatase	Meso Scale Discovery
	Sclerostin	Meso Scale Discovery
	Lipopolysaccharide binding protein	Meso Scale Discovery
	N-terminal pro–brain natriuretic peptide	MilliPlex
	Growth hormone	MilliPlex
	Pregnancy-associated plasma protein A	Alpco

Note: Meso Scale Discovery, Inc., Gaithersburg, MD; Millipore Inc., Billerica, MA; diaDexus, Inc., South San Francisco, CA; American Diagnostics, Hauppage, NY; Assay Pro, St. Charles, MO.