






ORIGINAL RESEARCH

Validation of an Integrated Genetic-Epigenetic Test for the Assessment of Coronary Heart Disease

Robert Philibert , MD, PhD; Timur K. Dogan, PhD; Stacey Knight , PhD; Ferhaan Ahmad , MD, PhD; Stanley Lau , MD; George Miles, MD, PhD; Kirk U. Knowlton , MD; Meeshanthini V. Dogan, PhD

BACKGROUND: Coronary heart disease (CHD) is the leading cause of death in the world. Unfortunately, many of the key diagnostic tools for CHD are insensitive, invasive, and costly; require significant specialized infrastructure investments; and do not provide information to guide postdiagnosis therapy. In prior work using data from the Framingham Heart Study, we provided in silico evidence that integrated genetic–epigenetic tools may provide a new avenue for assessing CHD.

METHODS AND RESULTS: In this communication, we use an improved machine learning approach and data from 2 additional cohorts, totaling 449 cases and 2067 controls, to develop a better model for ascertaining symptomatic CHD. Using the DNA from the 2 new cohorts, we translate and validate the in silico findings into an artificial intelligence–guided, clinically implementable method that uses input from 6 methylation-sensitive digital polymerase chain reaction and 10 genotyping assays. Using this method, the overall average area under the curve, sensitivity, and specificity in the 3 test cohorts is 82%, 79%, and 76%, respectively. Analysis of targeted cytosine-phospho-guanine loci shows that they map to key risk pathways involved in atherosclerosis that suggest specific therapeutic approaches.

CONCLUSIONS: We conclude that this scalable integrated genetic–epigenetic approach is useful for the diagnosis of symptomatic CHD, performs favorably as compared with many existing methods, and may provide personalized insight to CHD therapy. Furthermore, given the dynamic nature of DNA methylation and the ease of methylation-sensitive digital polymerase chain reaction methodologies, these findings may pave a pathway for precision epigenetic approaches for monitoring CHD treatment response.

Key Words: artificial intelligence ■ coronary heart disease ■ diagnosis ■ epigenetics ■ genetics ■ machine learning ■ methylation-sensitive digital PCR

Coronary heart disease (CHD) is the leading cause of death globally, including in the United States, with 9 million people dying annually from myocardial infarctions.¹ More than 90% of the risk for these events is attributable to modifiable factors.² To improve prevention and achieve greater levels of survival of cardiac events, there is a need for more sensitive and scalable methods to detect both the early risk and the current presence of CHD. In previous

work, we have shown that artificial intelligence (AI)-aided diagnostic approaches incorporating epigenetic (DNA methylation) effects and gene-methylation interactions (GxMeth) can be translated into clinically implementable methods that can outperform standard lipid-based approaches in predicting 3-year risk for the onset of symptomatic CHD.³ However, whether clinical translation could be successfully achieved for similar in silico findings⁴ for the detection of stable CHD has

Correspondence to: Robert Philibert, MD, PhD, Cardio Diagnostics Inc., 311 West Superior Street, Suite 444, Chicago, IL 60654. Email: rphilibert@cardiodiagnosticsinc.com

This manuscript was sent to Daniel T. Eitzman, MD, Senior Guest Editor, for review by expert referees, editorial decision, and final disposition.

Supplemental Material is available at <https://www.ahajournals.org/doi/suppl/10.1161/JAHA.123.030934>

For Sources of Funding and Disclosures, see page 11.

© 2023 The Authors and Cardio Diagnostics, Inc. Published on behalf of the American Heart Association, Inc., by Wiley. This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

JAHA is available at: www.ahajournals.org/journal/jaha

CLINICAL PERSPECTIVE

What Is New?

- Prior studies have shown that genetic and epigenetic variation are associated with coronary heart disease risk factors and coronary heart disease itself.
- This study extends prior studies to show that an artificial intelligence-guided method that simultaneously considers both genetic and DNA methylation information can be used to assess current coronary heart disease status in a clinically meaningful manner.

What Are the Clinical Implications?

- This form of epigenetic testing may provide a rapid, scalable, and sensitive alternative to currently established methods for assessing current CHD status.
- Because the 6 methylation indices are both dynamic and map to molecular pathways of established clinical risk factors, it may be possible that the epigenetic information contained in the test could be used to inform treatment choice and guide the evaluation of intervention effectiveness.
- Future studies to test this hypothesis are indicated.

Nonstandard Abbreviations and Acronyms

CpG	cytosine-phospho-guanine
CXCL1	C-X-C motif chemokine ligand 1
DHCR24	24-dehydrocholesterol reductase
EHD4	EH domain containing 4
FHS	Framingham Heart Study
GxMeth	gene-methylation interactions
IM	Intermountain Healthcare
INOCA	ischemia with non-obstructive coronary arteries
INSPIRE	Intermountain Healthcare Biological Samples Collection Project and Investigational Registry for the On-Going Study of Disease Origin, Progression, and Treatment
MSdPCR	methylation-sensitive digital polymerase chain reaction
TMEM18	transmembrane protein 18
UTR	untranslated region

remained uncertain. Instead, current guidelines for assessing the presence of stable CHD stipulate the use of established measures such as the use of nonemergent

exercise stress tests with ECGs or coronary computed tomographic angiography, depending on factors such as the degree of clinical suspicion, the ability to exercise, and the likelihood of high-risk events.^{5,6}

The performance of these more established methods for assessing stable CHD status has been extensively examined. The most commonly used and least invasive option, exercise ECG, has a sensitivity of only 58%.⁷ In contrast, the best-performing method, coronary computed tomographic angiography, is reported to have a sensitivity of 97%.⁸ However, coronary computed tomographic angiography is expensive, requires the use of contrast dye, entails significant exposure to ionizing radiation, and is associated with an ≈1% rate of serious morbidity and occasional death.^{9,10} Critically, the sensitivity and the specificity of these and the other methods are usually calculated against invasive coronary angiography as a reference. However, while angiography is the gold standard for detecting obstructive CHD, it does not detect the presence of myocardial ischemia with nonobstructive coronary artery disease (INOCA), a significant cause of CHD mortality.¹¹ Therefore, since up to 65% of women and 32% of men undergoing angiography for suspicion of CHD with stable angina have no significant obstruction,¹² it is perhaps more accurate to state that our understanding of the performance of each of these tests for the overall syndrome of stable CHD, and not just obstructive epicardial coronary artery disease, is still being refined.¹¹

Limitations in sensitivity are not the only challenge to the use of these methods in the clinical setting. In managed health care settings, access to these tests may be difficult because of cost, ranging from a minimum of \$1125 for an exercise ECG¹³ to a maximum of \$9224 for a cardiac positron emission tomography scan.^{14,15} Furthermore, because many of these tests typically require access to specialized equipment or health care providers, congested testing schedules can result in further delays.¹⁶ As a result of these and other structural health care barriers, patients often do not receive timely and affordable cardiac assessments that they need.

Recently, advancements at the intersection of high-throughput epigenetics and AI have led to the development of AI-driven molecular technologies that may provide a more scalable and readily accessible method for assessing stable CHD. CHD results from a combination of both environmental and genetic factors.² The overall genetic contribution to CHD is relatively small, with only 16% of the population-attributable risk being secondary to common genetic variation.¹⁷ Instead, the vast majority of risk is acquired from environmental stressors or lifestyle factors such as a diet rich in saturated fats, smoking, and sedentary lifestyle. Many of these risk factors have their own unique DNA methylation signatures in peripheral whole blood, often

at key loci being affected by interactions between local or distant genetic variation (GxMeth effects). Therefore, we previously hypothesized that an AI-based approach that considered the impact of genetics, DNA methylation, and GxMeth (genetic \times DNA methylation) effects simultaneously would enable a highly sensitive approach for predicting CHD status. In 1545 subjects in the FHS (Framingham Heart Study), we used a machine learning approach to generate a random forest classifier that incorporated the methylation information from 4 cytosine-phospho-guanine (CpG) sites, 2 single nucleotide polymorphisms (SNPs), age, and sex to predict symptomatic CHD with an accuracy, sensitivity, and specificity of 78%, 75%, and 80%, respectively.⁴

Our study was subject to several limitations. First, the number of subjects with CHD in the FHS cohort was relatively small. Second, the methylation information used in our analyses was acquired from genome-wide arrays, which are time consuming to process, costly, and relatively inaccurate.¹⁸ Third, the FHS study participants were all White individuals and from the northeastern United States.¹⁹ Fourth, the FHS CHD assessments represented only best estimates from the Framingham Endpoint Review Committee.¹⁹ Diagnostic testing was not performed as part of the study protocol. It was uncertain whether our analysis and results would be applicable to subjects ascertained by other clinical methods, recruited in other regions of the country, or being of other race or ethnicity.

In this study, we develop and validate an integrated genetic–epigenetic approach for the detection of CHD. We include data from 3 independent cohorts ascertained in differing regions of the country and translate the significant DNA methylation findings into accurate, precise methylation-sensitive digital polymerase chain reaction (MSdPCR) assays. The study's end product comprises a panel of MSdPCR and genotyping assays, along with a validated machine learning algorithm. Together, these components form a potentially clinically implementable tool for predicting CHD status.

METHODS

Data Availability

Information for obtaining the use of data from the FHS is available from the National Library of Medicine Database of Genotypes and Phenotypes. The data from the Intermountain Healthcare (IM) and University of Iowa cohorts to support the findings of this study are available from Dr. Meeshanthini Dogan upon reasonable request and subject to commercial restrictions.

Study Cohorts

Data were obtained from 3 repositories, the FHS, the IM Heart Institute INSPIRE (Intermountain

Healthcare Biological Samples Collection Project and Investigational Registry for the On-Going Study of Disease Origin, Progression, and Treatment) registry, and the Iowa CHD Repository (Iowa). A full description of the procedures and protocols used in the FHS (the National Library of Medicine Database of Genotypes and Phenotypes study accession: phs000007) has been provided elsewhere.^{20,21} Following written consent to participate in genetics research, subjects underwent phlebotomy. The CHD status of these individuals was assessed at each exam cycle, using data from the clinical intake as well as medical records, by the Framingham Endpoint Review Committee.¹⁹ The data included in this study are from the Offspring cohort who attended the eighth examination cycle, which was conducted between 2005 and 2008, consented to genetics research, and have peripheral blood genome-wide DNA methylation data. All clinical and biological data for the FHS Offspring cohort, including self-reported subject sex, were obtained in a fully anonymized form through the National Library of Medicine Database of Genotypes and Phenotypes (<https://dbgap.ncbi.nlm.nih.gov>). The University of Iowa Institutional Review Board approved all analyses described in this study (IRB No. 20150302).

The design and procedures of the IM registry have been previously described.^{22,23} In brief, IM patients who underwent clinically indicated coronary angiography were solicited to participate in the registry. All subjects provided informed consent to participate in the registry. However, only those CHD case and control subjects who provided DNA at the time of catheterization were eligible to participate in this study. A CHD case subject was defined as an adult aged >18 years who did not have a history of CHD or myocardial infarction before the index coronary angiogram but had a clinical diagnosis of CHD ($>70\%$ stenosis) on angiography. A control subject was defined as an adult aged >18 years who did not have a history of CHD or myocardial infarction before the index coronary angiogram and had no clinical diagnosis of obstructive coronary artery disease ($<50\%$ stenosis) at the index coronary angiography. Whole blood DNA was collected, and deidentified DNA and clinical data, including self-reported subject sex, were supplied to Cardio Diagnostics for further analyses. The procedures and protocols used for the analyses of the IM materials were approved by the IM Institutional Review Board (IRB No. 1024811).

Iowa CHD subjects were individuals aged >18 years who presented to the University of Iowa Hospital and Clinics for evaluation of CHD and were hospitalized. After providing full informed consent, each subject was interviewed to confirm the CHD presentation history, and then underwent phlebotomy. The clinical diagnosis of CHD was confirmed from the discharge medical summary. Control subjects without symptomatic CHD

were recruited by advertisements within the University of Iowa community. After informed consent was obtained from each of these individuals, they were interviewed with a variety of instruments including those from the PhenX toolkit and then underwent phlebotomy.²⁴ Subjects' report of an absence of CHD was confirmed using hospital records, when available. Vital signs, hemoglobin A_{1c}, and cholesterol levels, when available, are from nonacute cardiac encounters in the prior year. Subject sex was as per subject self-report. DNA from cases and controls was processed from whole blood using standard procedures.²⁵ All procedures and protocols used in this registry were approved by the University Iowa Institutional Review Board (IRB No. 201910834).

Genome-Wide Genetic and Epigenetic Data

This study used existing genome-wide genetic and epigenetic data from the FHS. The DNA methylation data were obtained using the Infinium HumanMethylation450 BeadChip array (Illumina, San Diego, CA). In total, data for 2567 subjects were available from the eighth examination cycle. As a first step, we performed standard sample- and probe-level quality control as described in previous studies, which resulted in the retainment of DNA methylation data from 2560 samples at 403192 loci.^{4,26–29} Genome-wide genotype data obtained using the GeneChip HumanMapping 500K array (Affymetrix, Santa Clara, CA) were available for 2406 of the remaining 2560 samples. After standard sample- and probe-level quality control procedures using PLINK were performed. This resulted in the removal of 111 samples, leaving the total number of samples and SNPs remaining at 2295 and 472 822, respectively.^{4,29,30} Finally, because there is the potential for interrelatedness of some of the FHS subjects, the genetic data were subjected to relatedness analysis in PLINK. In total, data from 696 subjects were removed secondary to genetic relatedness (identity by state >0.1875, which is halfway in between second- and third-degree relatives). After removal of subjects for lack of clinical or molecular data, complete information for 2111 FHS subjects were available.

Genome-wide DNA methylation and genetic assessments for the IM cohort were conducted by the University of Minnesota Genome Center using the Infinium MethylationEpic Beadchip array and the Infinium Multi-Ethnic Global BeadChip array (Illumina), respectively. These data were then subjected to the same quality control procedure as the FHS cohort.³ In total, 862 593 methylation and 818 046 SNP loci survived quality control measures. For the data mining analyses, we retained loci common to both the Illumina Infinium HumanMethylation450 BeadChip

array and Illumina Infinium MethylationEpic Beadchip array arrays, resulting in 437 242 loci being available for further analysis. Similarly for SNPs, we retained those common to both genotyping arrays, resulting in 80 371 loci being available for further analysis. After removal of subjects for incomplete data, including MSdPCR data, information for 245 subjects was available.

AI Model and Validation of Epigenetic and Genetic Biomarkers

Figure 1 outlines the approach used to develop and validate the integrated genetic–epigenetic machine learning model for CHD prediction. The process involved several preliminary steps: (1) quality control of genetic and epigenetic data, as detailed in the preceding section; (2) subsetting markers shared by the FHS and the IM cohorts; and (3) assessment of methylation markers for translation. The primary goal of this study was to develop a clinically implementable integrated genetic–epigenetic test that consists of standard fluorescent genotyping assays for SNP assessment, stand-alone MSdPCR assays for methylation assessment, and a machine learning prediction model for CHD status prediction. To maximize the robustness of the MSdPCR assays, data mining was restricted to methylation loci exhibiting a $\Delta\beta$ value exceeding 0.01. Additionally, the analysis was confined to markers that were common between the FHS and IM cohorts. After completing the preliminary steps, a total of 10 484 methylation markers and 67 749 SNP markers were used in data mining. As a final preparatory step, the FHS cohort was divided into a training set (75%) and a test set (25%). The training set contained data from 183 CHD subjects and 1400 controls, while the test set contained 61 CHD subjects and 467 controls.

Data mining was conducted exclusively on the FHS training set. Our mining strategy incorporated random undersampling with replacement to equalize class representation and bolster robustness. This undersampling process was executed 1000 times using random forest, logistic regression, and support vector algorithms, and the resulting area under the receiver operating characteristic curve (AUC) scores were averaged to gauge predictive performance. The scikit-learn machine learning library facilitated algorithm implementation, while mpi4py enabled parallelization.^{31,32} In the initial phase, SNP and methylation markers with AUC values <0.5 were eliminated. Subsequently, random feature selection was used to identify integrated genetic–epigenetic markers capable of predicting CHD status. The data mining was performed on a 280-core high computing system for \approx 3 months.

The algorithm development process involved the use of digital methylation assessments of DNA samples already profiled with arrays to renormalize the

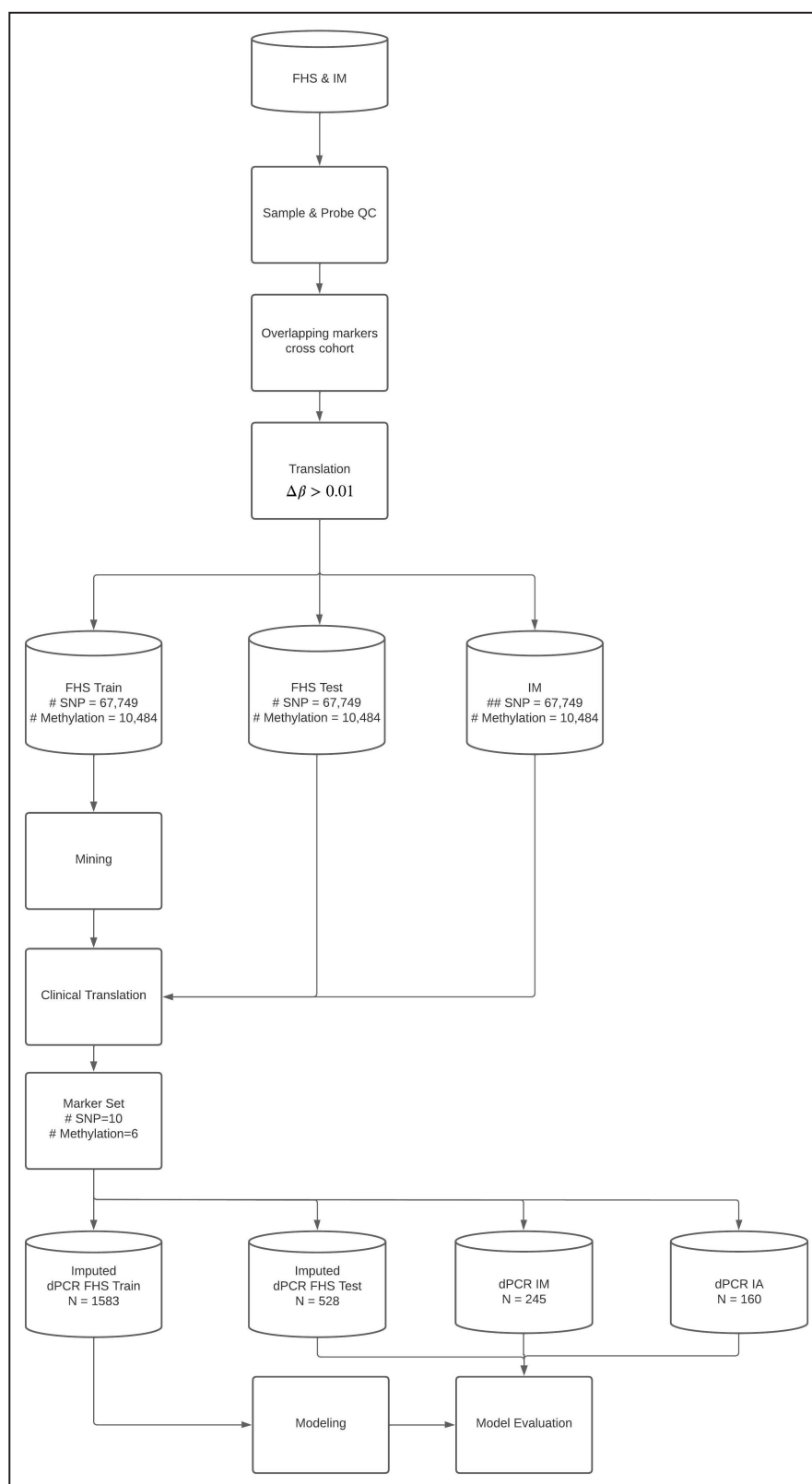


Figure 1. Flowchart summarizing steps used in data cleaning, formation of test and training sets, data mining, initial algorithm development, tuning of algorithm using MSdPCR data, and final model evaluation.

dPCR indicates digital polymerase chain reaction; FHS, Framingham Heart Study; IA, Iowa; IM, Intermountain Healthcare; QC, quality control; and SNP, single nucleotide polymorphism.

distribution of the array data at select loci. This approach was advantageous because MSdPCR is a reference-independent method for assessing methylation as opposed to the complex, error-prone reference dependent method used to normalize methylation signal at each locus. The methylation probes from the top performing sets were then used to construct MSdPCR assays. After the successful translation of a probe set highly predictive of CHD, the MSdPCR markers were then used to assess methylation status of the entire IM cohort. Support vector machine was used to develop an imputation model between array-based and MSdPCR-based methylation values. An exhaustive feature selection step was then performed on the imputed FHS training set. As a result of the data mining, the best-performing feature set comprised 6 methylation and 10 SNP markers, which translated into standard MSdPCR assays and hydrolyzable fluorescent primer probe genotyping assays (Data S1).

Using the imputed MSdPCR methylation data and SNP data, a balanced bagging classifier model^{33–39} from the Imbalanced-Learn Python toolbox⁴⁰ was trained on the FHS training set. This final CHD status prediction model was then tested in the FHS test set, and externally validated in the IM and Iowa cohorts.

MSdPCR assessments of DNA methylation in subjects from the IM and Iowa cohorts were conducted using our standard methods.^{3,41,42} In brief, 1 µg of DNA underwent bisulfite conversion using the EpiTect Bisulfite kit (Qiagen, Hilden, Germany) according to manufacturer's directions and eluted in 70-µL volume. Fourteen cycles of high stringency polymerase chain reaction amplification of the target region were then performed on a 3-µL aliquot of each sample using a set of amplicon-specific proprietary primers. Finally, an aliquot of the enriched amplicon target solution was diluted 1:1500, mixed together with primer and probes specific for the targeted loci and droplet digital polymerase chain reaction reagents, partitioned into droplets with a droplet generator (Bio-Rad Laboratories, Hercules, CA), and then polymerase chain reaction amplified. The methylation status of each droplet was then determined using a Bio-Rad QX-200 Reader and the percentage of methylation status of each sample was imputed using the Bio-Rad QuantaSoft software.

RESULTS

Study Cohorts

The clinical and demographic characteristics of the 3 cohorts included in this study are presented in Table 1. The first cohort of subjects is from the well-known FHS Offspring cohort. After excluding subjects due to factors such as incomplete genome-wide data or

consanguinity, a total of 2111 FHS subjects remained, 244 of them diagnosed with CHD and 1867 without a current diagnosis of symptomatic CHD. All FHS subjects were from the Northeast United States and were exclusively White individuals. While the majority of the FHS subjects were women (1163/2111 subjects; 55%), the majority of those with CHD were men (164/244 subjects; 67%). On average, CHD subjects were 5 to 6 years older than those without CHD. Those with a CHD diagnosis at the 2008 intake wave had lower cholesterol levels than the controls and were more likely to be on statins (both $P < 0.001$).⁴ Women had both higher total cholesterol and high-density lipoprotein than their male counterparts ($P < 0.001$). Conversely, no significant differences in hemoglobin A_{1c} levels were observed between male and female subjects ($P > 0.65$).

The IM cohort consisted of 245 subjects, of whom 113 (46%) were men and 132 (54%) were female. The men were on average 7 years younger than the women ($P < 0.001$), and CHD subjects were on average 2 to 3 years younger than those without CHD. While total serum cholesterol levels were similar to those of the FHS subjects, high-density lipoprotein levels were significantly lower ($P < 0.001$). Additionally, both systolic and diastolic blood pressures were significantly higher (both $P < 0.001$) than those in the FHS cohort.

In the Iowa cohort, similar to the IM cohort, subjects with CHD tended to be in their mid-60s. Also, like the FHS cohort, a majority of those with CHD were men (56/76 subjects; 72%). Among the 76 CHD cases, 70 were White individuals, 1 Black individual, 1 Hispanic individual, 2 Native American individuals, and 2 were of unknown race and ethnicity. Similarly, almost all the controls were also White individuals, except for 3 Asian individuals, 1 Black individuals, and 1 of unknown race and ethnicity. Because the subjects with CHD were not systematically assessed before hospitalization, rigorous standardized laboratory and vital sign assessments for these case subjects are not available. For the controls, who tended to be significantly younger than the cases, systematic vital sign and laboratory assessments were available and are provided in Table 1.

Genetic and Epigenetics Biomarkers

Using genome-wide genetic and epigenetic data from subjects in the FHS training set, we developed a prediction model for current CHD. The most robust model consisted of 6 methylation sites (cg03725309, cg12586707, cg04988978, cg17901584, cg21161138, and cg12655112) and 10 SNPs (rs710987, rs1333048, rs12129789, rs942317, rs1441433, rs2869675, rs4639796, rs4376434, rs12714414, and rs7585056) for CHD prediction (see Table 2).

Whereas rapid, accurate stand-alone fluorescent genotyping (eg, Taqman) assays are readily available for

Table 1. Summary of Demographics and Conventional Coronary Heart Disease Risk Factors for FHS Offspring Cohort Training and Test Set, Iowa Validation Set, and IM Validation Set

	FHS training (n=1583)		FHS test (n=528)		Iowa validation (n=160)		IM validation (n=245)	
	No CHD	CHD	No CHD	CHD	No CHD	CHD	No CHD	CHD
Sex, n								
Female	812	60	271	20	52	20	63	69
Male	588	123	196	41	32	56	53	60
Age, y								
Female	65.8±8.9	73.2±8.8	66.0±8.4	71.8±8.8	47.2±10.2	61.8±14.3	65.9±14.0	69.6±11.9
Male	64.9±8.7	70.0±7.9	64.8±8.5	70.1±7.9	46.3±11.8	64.9±10.2	59.7±16.5	62.0±15.6
Total cholesterol, mg/dL								
Female	197.6±35.1	173.6±35.5	200.6±36.3	176.4±40.0	188.2±54.8	153±21.2	180.7±35.6	190.0±50.5
Male	179.1±32.5	151.0±32.9	176.8±31.9	147.0±24.4	172.4±30.7	171.3±41.9	171.2±35.9	168.6±48.8
High-density lipoprotein cholesterol, mg/dL								
Female	64.8±18.6	60.0±18.8	66.1±19.2	58.1±12.8	70.1±20.1	53.3±21.1	55.6±20.1	45.4±15.3
Male	50.5±13.9	45.6±11.4	49.9±14.6	45.0±11.1	48.3±16.7	51.7±20.1	38.5±11.7	38.6±10.5
HbA _{1c} , %								
Female	5.7±0.6	6.1±0.9	5.6±0.4	6.0±0.9	5.3±0.4	8.0±1.8	6.1±1.4	7.1±1.0
Male	5.7±0.7	6.0±0.9	5.7±0.8	6.0±1.0	5.6±0.3	6.8±1.3	6.1±0.7	6.2±1.2
Systolic blood pressure, mmHg								
Female	127.7±17.6	135.8±16.7	129.0±18.4	132.1±19.8	120.9±16.6	133.2±24.2	151.1±22.0	151.9±27.5
Male	129.6±16.7	124.4±18.3	128.4±16.5	133.1±19.4	129.7±11.8	140.2±19.9	141.6±24.0	140.1±20.4
Diastolic blood pressure, mmHg								
Female	72.7±9.9	69.2±10.9	73.6±10.2	67.8±12.0	79.5±11.5	72.4±18.6	80.0±12.0	78.6±14.4
Male	76.5±10.3	68.8±10.9	75.9±10.4	70.3±13.4	89.3±9.8	81.6±15.5	84.0±14.5	81.3±11.4

CHD indicates coronary heart disease; FHS, Framingham Heart Study; HbA_{1c}, hemoglobin A_{1c}; HDL, high-density lipoprotein; and IM, Intermountain Healthcare.

Table 2. List of Cytosine-Phospho-Guanine and Single Nucleotide Polymorphism Sites Used in the Model

Motif	Chromosome	Gene	Gene location
cg03725309	1	SARS1	Body
cg12586707	5	CXCL1	3' Intergenic region
cg04988978	17	MPO	5' Promoter region
cg17901584	1	DHCR24-DT	Gene body
cg21161138	5	AHRR	Gene body
cg12655112	15	EHD4	Gene body
rs710987	5	LINC010019	Gene body
rs1333048	9	CDKN2B	3' Intergenic region
rs12129789	1	KCND3	Gene body
rs942317	14	KTNN1-AS1	Gene body
rs1441433	4	PPP3CA	Gene body
rs2869675	20	PREX1	Gene body
rs4639796	1	ZBTB41	Gene body
rs4376434	7		Intergenic region near LINC00972
rs12714414	2		Intergenic region near TMEM18
rs7585056	2		Intergenic region near TMEM18

Locations and Human Genome Organization gene abbreviations per University of California, Santa Cruz Genome Browser using hg38 build. Motifs were assigned to a gene if they were within 5 kb of the nearest exon of the listed gene. LINC00972 indicates long intergenic non-protein coding RNA 972; TMEM18, transmembrane protein 18.

SNPs, similar tools for assessing DNA methylation at select loci are not yet commercially available. Nevertheless, assays for DNA methylation are critical to the cost-effective, stand-alone clinical implementation of the test. Therefore, MSdPCR assays were devised for measuring methylation values at each of our 6 CpG sites.

Figure 2 shows the relationship between methylation values measured using the Illumina array (x axis) as compared with those using MSdPCR (y axis) for subjects in the IM cohort. The methylation values for 4 of the 6 MSdPCR assays demonstrate high linear correlations ($r \geq 0.85$) with their corresponding values derived from the Illumina array. One MSdPCR assay (Dcg12655112) has good ($r=0.76$) linear correlation with its corresponding (cg12655112) array values and the last MSdPCR assay (Dcg12586707) has a moderate correlation ($r=0.55$) with its corresponding (cg12586707) array values. Consistent with prior results, the range for the MSdPCR assay was equivalent or greater than the values for the array values.⁴¹

Overall and Sex-Based Performance of Integrated Genetic-Epigenetic Biomarkers for CHD Status Prediction

The performance of the CHD detection model, its AUC, and its sensitivity and specificity by sex are provided in Table 3 for the FHS test set (nonindependent test set) and the IM (first independent, external validation set)

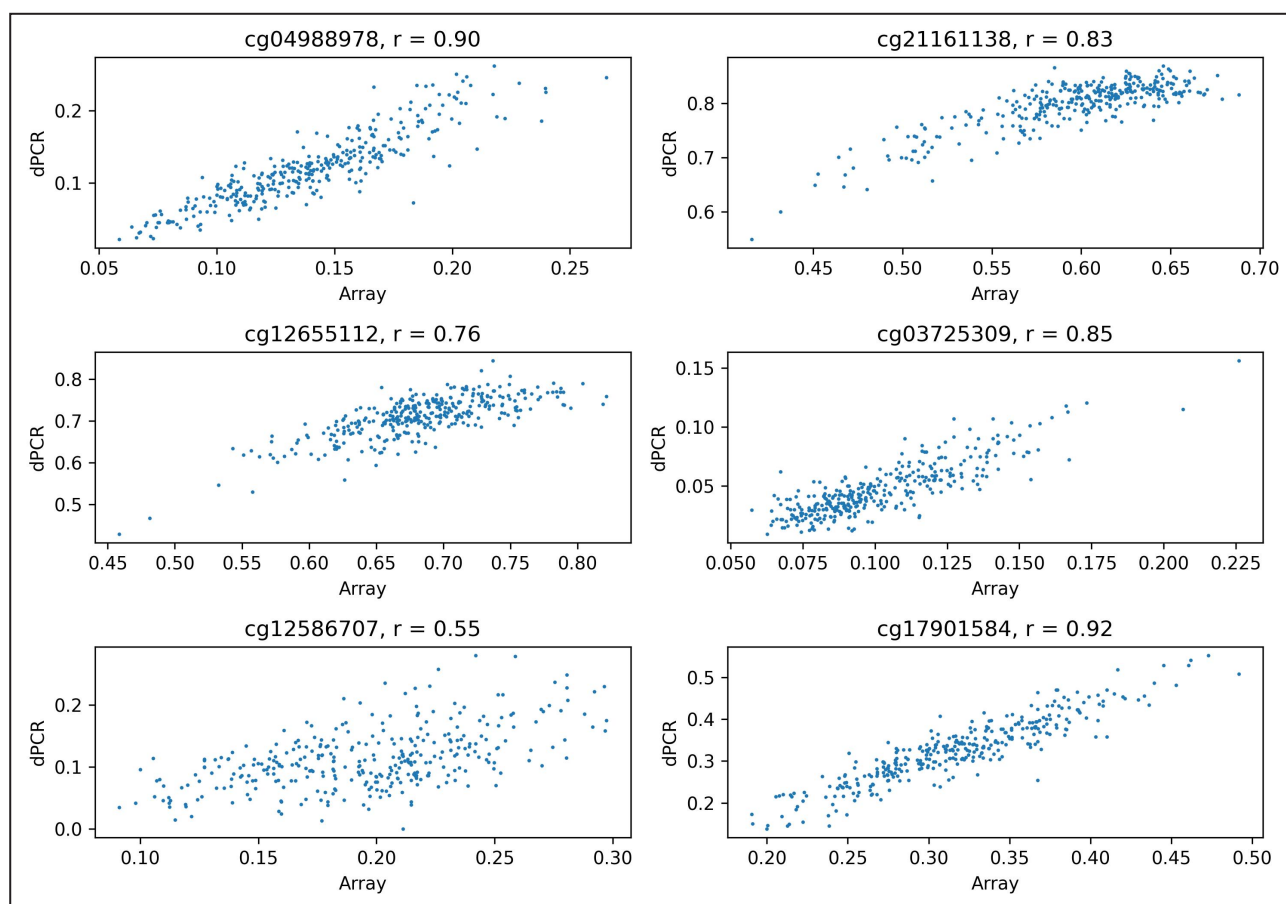


Figure 2. The relationship between the MSdPCR- and Illumina array-derived methylation values for each of the subjects in the IM cohort.

Methylation is expressed as fractional methylation (ie, between 0 and 1) with the linear relationship between each of the values being expressed as a Pearson correlation coefficient (r). dPCR indicates digital polymerase chain reaction; IM, Intermountain Healthcare; and MSdPCR, methylation-sensitive digital polymerase chain reaction.

and the Iowa cohorts (second independent, external validation set). Overall, the model performed very well with AUC, sensitivity, and specificity of 0.82, 0.78 and 0.74, respectively, in the FHS test cohort. The performance was highly generalizable to the external validation cohorts; in the IM cohort, the AUC, sensitivity, and specificity were 0.75, 0.76 and 0.71, respectively, and in the Iowa cohort, AUC, sensitivity, and specificity were 0.88, 0.82 and 0.82, respectively.

Table 4 provides the methylation values for the 6 MSdPCR assays (designated Dcg) used in our algorithm in the case and control subjects from the Iowa cohort. Interestingly, in each case, the subjects with CHD had significantly lower methylation values than the control subjects (all $P < 0.0001$).

DISCUSSION

This study presents a clinically implementable integrated genetic–epigenetic test guided by AI for CHD.

A major strength of this study is the inclusion of 3 independent cohorts with differing definitions of CHD. Another major strength of this study lies in the use of AI to identify molecular genetic, epigenetic, and GxMeth patterns associated with CHD without relying on a priori assumptions or solely considering biomarkers previously linked to conventional risk factors such as lipids or diabetes. This approach aims to minimize bias from consideration of only a few risk factors to uncover nonlinear signals associated with CHD and potentially to elucidate new combinations of pathways involved in the pathogenesis of CHD.

With this approach, 6 DNA methylation markers were incorporated that map to at least 6 distinct, potentially modifiable pathways known to be involved in the pathogenesis of ischemic heart disease into the algorithm. The first pathway, cholesterol biosynthesis, is notable due to the importance of statins in primary and secondary prevention of CHD. The CpG site that maps to this pathway is cg17901584, which is found in an intron of 24-dehydrocholesterol reductase

Table 3. Performance of Precision CHD in the FHS, IM, and Iowa Cohorts

Cohort	AUC	Sensitivity	Specificity
FHS test			
Overall	0.82	0.78	0.74
Female	0.82	0.75	0.75
Male	0.81	0.80	0.72
IM			
Overall	0.75	0.76	0.71
Female	0.73	0.77	0.70
Male	0.77	0.75	0.72
Iowa			
Overall	0.88	0.82	0.82
Female	0.87	0.75	0.83
Male	0.83	0.84	0.81

AUC indicates area under the receiver operating characteristic curve; CHD, coronary heart disease; FHS, Framingham Heart Study; and IM, Intermountain Healthcare.

(DHCR24)–divergent transcript, a long noncoding RNA gene that is in a divergent (head-to-head) configuration with DHCR24, a key gene in cholesterol biosynthesis. Critically, because the intergenic region in both humans and avians is small (200 bp in humans) and studies of the orthologous genes in chickens have shown that gene expression is highly correlated,⁴³ it is likely that the 2 genes are coregulated by the same bidirectional promoter. At the same time, because CHD is associated with a decrease in methylation (demethylation) of cg17901584, and demethylation is associated with increased transcription, this suggests that CHD is associated with an increased level of transcription of DHCR24–divergent transcript relative to DHCR24. Understanding the consequences of any increased transcription on DHCR24 function and cholesterol biosynthesis will need further investigation for several reasons. First, long noncoding RNAs can alter their regulatory targets by several mechanisms at the transcriptional and posttranscriptional level.⁴⁴

Table 4. Methylation Values in Subjects With CHD and Control Subjects From Iowa

	Control	Case
MSdPCR assay	(n=84)	(n=76)
Dcg04988978, %	16.4±4.5	9.1±3.5
Dcg21161138, %	81.8±2.5	78.0±6.3
Dcg12655112, %	74.0±3.4	68.3±5.4
Dcg03725309, %	7.7±2.6	3.5±1.5
Dcg12586707, %	15.1±4.8	8.3±3.6
Dcg17901584, %	39.6±7.9	29.8±8.0

All *P* values <0.0001.

CHD indicates coronary heart disease; and MSdPCR, methylation-sensitive digital polymerase chain reaction.

Furthermore, unlike DHCR24, which has 9 well-defined exons,⁴⁵ DHCR24–divergent transcript has at least 2 transcripts of 2 and 3 exons that only partially overlap. Conceivably, each of these transcripts may have differing effects on DHCR24 gene function with cg17901584 demethylation differentially affecting their expression. Finally, demethylation can affect enhancer sites involved in long-distance chromatin interactions, and other key genes linked to atherosclerosis, such as proprotein convertase subtilisin/kexin type 9,⁴⁶ are less than 150 kb away. Systematic examinations using our more precise MSdPCR assays to determine whether DNA methylation directly predicts cholesterol levels and response to statin therapy are in order.

Similarly, the pathway tagged by cg04988978 highlights the complex relationships between the classic serological predictors of CHD. Cg04988978 maps to a CpG site 2 kb upstream of the first exon of myeloperoxidase, which is thought to contribute to atherosclerosis by oxidation of LDL.⁴⁷ Because demethylation of promoters is associated with increased gene transcription, it may be surmised that demethylation is positively correlated with myocardial infarctions due to the elevated level of myeloperoxidase. However, Fernández-Sanlés and colleagues⁴⁸ have shown that methylation of cg04988978 is positively associated with high-density lipoprotein cholesterol levels, and negatively associated with triglycerides and glucose levels. As each of these serum markers are independently associated with CHD, understanding the primacy and exact molecular mechanisms underlying the association of the cg04988978 locus with CHD may entail considerable additional investigation.

The third DNA methylation marker, cg21161138, maps to the aryl hydrocarbon receptor repressor. Demethylation at this locus is tightly correlated with demethylation of cg05575921, a marker of smoking intensity.^{49,50} Since smoking is a preventable cause of heart disease and reversion of aryl hydrocarbon receptor repressor methylation is a marker of smoking cessation, this marker could be used to not only help clinicians more exactly understand the pathogenesis of a given patient's CHD, but also could be reassessed as part of a more holistic precision epigenetic approach to smoking cessation and secondary CHD treatment.⁵¹

The DNA methylation marker cg03725309, which maps to candidate cis regulatory element in intron 1 of the seryl-tRNA synthetase 1 gene (*SARS1*), is significantly demethylated in CHD. Mutational analyses conducted in zebrafish demonstrated the essential role of this gene in normal vascular development.^{52,53} Demethylation at this locus is associated with obesity, coronary artery calcification, and cardiometabolic syndrome.^{54–56} Lower serine levels are not only associated with diabetes, but also supplementation of the diets of

diabetic mice with serine can reverse some of the pathology associated with diabetes.^{57,58} If these findings can be extended to CHD as well, serine supplementation, which has very few side effects, may be a new method through which to address the increasing societal burden of CHD, unlocking a new vista for personalized dietary-based approaches for CHD treatment and prevention.⁵⁹

The breadth of the gene networks regulating glucose metabolism and their profound impact on vulnerability to CHD is further emphasized by the additional value of the cg12655112 marker, which maps to intron 1 of the EH domain containing 4 (*EHD4*). *EHD4* is 1 of several human orthologues with high sequence identity that play key roles in the regulation of endocytic vesicles.⁶⁰ *EHD4* methylation has been negatively associated with serum glucose levels and *EHD4* expression predicts the success of pancreatic islet transplants.^{61–63} Finally, *EHD4* protein levels predict the development of diabetic cardiomyopathy in type 2 diabetic mice.⁶⁴

The importance of inflammation in the pathogenesis of CHD is highlighted by the contribution of the cg12586707 marker to CHD status prediction. Cg12586707 maps to a candidate cis regulatory element approximately 1.5 kb downstream of the 5' UTR of the C-X-C motif chemokine ligand 1 (*CXCL1*) gene. *CXCL1* is a key member of a group of chemotactic messengers involved in the pathogenesis of a number of inflammatory disorders and has an important role in the regulation of angiogenesis.⁶⁵ The CpG site sits in a large peak of H3K27 acetylation activity downstream of the 3' UTR that is continuous with the H3K27 signal from the *CXCL1* gene proper. This is important because prior studies have shown that the 3' UTR region of *CXCL1* is particularly critical in maintaining mRNA stability.⁶⁶ Because 3' UTR splicing can be variable, and this variability can affect the stability and spatiotemporal timing of translation,⁶⁷ cg12586707 methylation status may reflect *CXCL1* mRNA stability that is important in the pathogenesis of CHD. Because the administration of aspirin increases *CXCL1* expression,⁶⁸ and because aspirin has a mixed track record as a therapeutic agent in heart disease,^{69,70} the MSdPCR assay for this locus may identify a subgroup of patients who preferentially respond to aspirin therapy.

Although most of the power and the sensitivity of the CHD status prediction model is driven by the DNA methylation markers, the SNPs make noticeable contributions to the specificity through their GxMeth interaction effects. The potential molecular mechanisms behind these effects are often complex, with very few GxMeth interaction effects having been elucidated. Two of the SNPs, rs7585056 (CHR2:631528) and rs12714414 (CHR2:651408), map within 20 kb of one another to the intergenic region just downstream of the longest splice variant of transmembrane protein

18 (*TMEM18*). This is particularly interesting because genetic variation in *TMEM18* is well established as being associated with obesity, type 2 diabetes and CHD.^{71,72} Since rs7585056 and rs12714414 are outside the coding region of *TMEM18* and typical haplotype blocks in humans are on the order of only 5 to 10 kb, it is unlikely that these 2 SNPs are tagging genetic variation directly in the coding region of *TMEM18*. However, because there are 3 different splicing variants of the 3' UTR, their location distal to the 3' UTR may suggest that these SNPs are tagging variation that alters the number, ratio, or stability of *TMEM18* transcripts.

Although there are numerous models for predicting incident CHD,⁷³ to our knowledge, this is the first integrated genetic–epigenetic algorithm for current CHD status. In 2022, Zhang and colleagues⁷⁴ described 3 different machine learning models that used genome-wide methylation and expression information to predict current CHD status in the FHS. The performance metrics of each of the models were similar to the current values, but there was no external validation and no translation of either the methylation or expression assessments into a clinically implementable format.

An important feature of this panel of biomarkers and its corresponding CHD status prediction algorithm is the breadth of sensitivity across three very differently ascertained cohorts of CHD subjects. Because of the manner in which they were collected, the FHS subjects have perhaps the greatest diversity in severity and breadth of presentation. Since the diagnosis of CHD could be established by the physician review board solely on signs and symptoms reported to the research team without any additional testing, it is likely that many of the subjects who participated in this community study may have had mild CHD. In contrast, the Iowa cohort has the greatest severity of disease. Each of the Iowa subjects were admitted to the hospital for ischemic heart disease, with almost all having had a myocardial infarction, and some not surviving to discharge. Similarly, the FHS cohort also likely had the greatest breadth of CHD pathophysiology. As discussed above, INOCA is difficult to diagnose and may be more frequent in women.⁷⁵ INOCA thought by some to be a factor in the underrecognition of CHD in Black women.⁷⁶ Because the FHS is a community sample, it will have a mixture of both obstructive and nonobstructive CHD. In contrast, all the case subjects from the IM cohort were diagnosed with obstructive CHD. Nevertheless, despite the breadth of type and severity of presentations in these 3 cohorts, the algorithm performed with an average sensitivity of 80% and 76% for men and women, respectively, across the FHS and the IM and Iowa cohorts. This compares very favorably with that of conventional exercise treadmill testing, whose sensitivity is only 45% to 68%, but unfavorably

to coronary computed tomographic angiography, whose sensitivity is $\approx 97\%$, for the detection of obstructive CHD, depending on the source cited,⁷ while the sensitivity of each of the methods for nonobstructive forms of CHD, at this time, remains unknown.

Even though the model performed well for subjects with obstructive and nonobstructive CHD, we believe that this model can be further optimized in the future using a larger cohort of even better characterized subjects. For instance, those with INOCA in the IM cohort may have been mistakenly misclassified as unaffected in this study because the definition of CHD for this cohort focused only on obstructive CHD on the basis of angiography.^{77,78} Consistent with this assertion, in our 2021 study,³ we observed that 44 of the 159 of the IM subjects not found as having CHD at index ($<50\%$ stenosis) experienced a symptomatic CHD event within 3 years, including 19 who had an event within 6 months of angiography. Although not proof of INOCA, this cardiac morbidity and death in subjects with supposedly no CHD further highlights the need for a clinical tool capable of identifying those with obstructive and non-obstructive CHD in robustly characterized training and test cohorts.

Methodologically, our intentional selection of a wide range of CHD and the corresponding sensitivity achieved by the panel of 6 DNA methylation and 10 SNP markers will not only facilitate that additional optimization but may also facilitate the development of molecular endophenotypes predictive of key CHD features. For example, certain markers may distinguish between obstructive and nonobstructive presentations, allowing prioritization of those patients for angiography who may benefit from percutaneous coronary intervention and have improved clinical outcomes. Conversely, identifying those less likely to have obstructive CHD may avoid the expense and the potential complications of unnecessary angiograms.^{10,79} However, to further achieve these goals, it will be necessary to recruit and to characterize more subjects whose CHD is robustly characterized using clinical, functional, and diagnostic modalities.

Of note, our model does not include sex and age as predictive markers. Many current algorithms used in medicine, such as the atherosclerotic cardiovascular disease pooled cohort equation, are criticized for their overreliance on age as a predictor or their sex bias.^{80,81} To capture the biological variation predictive of CHD and avoid unintentional biases, our machine learning approaches did not include these variables in model development. Excluding age and sex did not affect the performance of the model, and, in fact, adding age and sex to the final model did not improve prediction. Furthermore, our model did not exhibit performance biases on the basis of sex or age. We are also cognizant that almost all the subjects used in the development of

the model are White individuals. Still, we note that genotype frequencies in the database of single nucleotide polymorphisms (<https://www.ncbi.nlm.nih.gov/snp/>) of the SNPs included in our model are informative in other ethnicities. Furthermore, the clustering of the non-White subjects from Iowa, whose data were used only for validation and were not used in model development, was undistinguishable from the clustering of the White subjects. Therefore, we are confident that our approach will generalize to other ethnicities and plan to demonstrate this in future studies.

In summary, in this study, we present the development and rigorous validation of a clinically implementable integrated genetic–epigenetic test guided by AI for the prediction of current CHD using biomaterials from 3 independent cohorts. We further discuss the targetable pathways the biomarkers map to and highlight the role of inflammation and serine metabolism in heart disease.

ARTICLE INFORMATION

Received May 10, 2023; accepted October 16, 2023.

Affiliations

Cardio Diagnostics Inc, Chicago, IL (R.P., T.K.D., M.V.D.); Department of Psychiatry (R.P.); and Department of Biomedical Engineering (R.P., M.V.D.), University of Iowa, Iowa City, IA Intermountain Heart Institute, Intermountain Healthcare, Salt Lake City, UT (S.K., K.U.K.); Department of Internal Medicine, University of Utah, Salt Lake City, UT (S.K.); Division of Cardiovascular Medicine, Department of Internal Medicine, University of Iowa, Iowa City, IA (F.A.); Southern California Heart Centers, San Gabriel, CA (S.L.); and Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX (G.M.).

Sources of Funding

This project was funded by Cardio Diagnostics Inc.

Disclosures

Drs. Philibert, T. Dogan, and M. Dogan are officers and stockholders of Cardio Diagnostics Inc. (www.cardiodiagnosticsinc.com). Drs. Philibert, M. Dogan, and Lau are members of the board of directors of Cardio Diagnostics Inc, and Dr Miles is a consultant for Cardio Diagnostics Inc. The use of DNA methylation and gene-methylation interaction effects for the assessment, diagnosis, and monitoring of cardiovascular disease is covered by US Patent 11 414 704, European Patent EP3472344B1, and other both granted and pending intellectual property claims elsewhere. The University of Iowa Research Foundation is entitled to royalties from use of this technology. The remaining authors have no disclosures to report.

Supplemental Material

Data S1

References 82–87

REFERENCES

1. Wang H, Naghavi M, Allen C, Barber RM, Bhutta ZA, Carter A, Casey DC, Charlson FJ, Chen AZ, Coates MM. Global, regional, and national life expectancy, all-cause mortality, and cause-specific mortality for 249 causes of death, 1980–2015: a systematic analysis for the global burden of disease study 2015. *Lancet*. 2016;388:1459–1544. doi: 10.1016/S0140-6736(16)31012-1
2. Ruff CT, Braunwald E. The evolving epidemiology of acute coronary syndromes. *Nat Rev Cardiol*. 2011;8:140–147. doi: 10.1038/nrcardio.2010.199

3. Dogan MV, Knight S, Dogan TK, Knowlton KU, Philibert R. External validation of integrated genetic-epigenetic biomarkers for predicting incident coronary heart disease. *Epigenomics*. 2021;13:1095–1112. doi: [10.2217/epi-2021-0123](https://doi.org/10.2217/epi-2021-0123)
4. Dogan MV, Grumbach IM, Michaelson JJ, Philibert RA. Integrated genetic and epigenetic prediction of coronary heart disease in the Framingham Heart Study. *PLOS One*. 2018;13:e0190549. doi: [10.1371/journal.pone.0190549](https://doi.org/10.1371/journal.pone.0190549)
5. Joseph J, Velasco A, Hage FG, Reyes E. Guidelines in review: comparison of esc and acc/aha guidelines for the diagnosis and management of patients with stable coronary artery disease. *J Nucl Cardiol*. 2018;25:509–515. doi: [10.1007/s12350-017-1055-0](https://doi.org/10.1007/s12350-017-1055-0)
6. Fihn SD, Gardin JM, Abrams J, Berra K, Blankenship JC, Dallas AP, Douglas PS, Foody JM, Gerber TC, Hinderliter AL. 2012 ACCF/AHA/ACP/AATS/PCNA/SCAI/STS guideline for the diagnosis and management of patients with stable ischemic heart disease: a report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines, and the American College of Physicians, American Association for Thoracic Surgery, Preventive Cardiovascular Nurses Association, Society for Cardiovascular Angiography and Interventions, and Society of Thoracic Surgeons. *Circulation*. 2012;126:e354–e471. doi: [10.1161/CIR.0b013e318277d6a0](https://doi.org/10.1161/CIR.0b013e318277d6a0)
7. Morrow D, De Lemos J. Chapter 40: Stable ischemic heart disease. In: Libby P, Bonow RO, Mann DL, Tomaselli GF, Bhatt D, Solomon SD, Braunwald E, eds. *Braunwald's Heart Disease: A Textbook of Cardiovascular Medicine*. 12th ed. Philadelphia, PA: Elsevier; 2022:739–785.
8. Gatti M, Gallone G, Poggi V, Bruno F, Serafini A, Depaoli A, De Filippo O, Conrotto F, Darvizeh F, Faletti R, et al. Diagnostic accuracy of coronary computed tomography angiography for the evaluation of obstructive coronary artery disease in patients referred for transcatheter aortic valve implantation: a systematic review and meta-analysis. *Eur Radiol*. 2022;32:5189–5200. doi: [10.1007/s00330-022-08603-y](https://doi.org/10.1007/s00330-022-08603-y)
9. Hausleiter J, Meyer T, Hermann F, Hadamitzky M, Krebs M, Gerber TC, McCollough C, Martinoff S, Kastrati A, Schömig A, et al. Estimated radiation dose associated with cardiac CT angiography. *JAMA*. 2009;301:500–507. doi: [10.1001/jama.2009.54](https://doi.org/10.1001/jama.2009.54)
10. West R, Ellis G, Brooks N. Complications of diagnostic cardiac catheterisation: results from a confidential inquiry into cardiac catheter complications. *Heart*. 2006;92:810–814. doi: [10.1136/hrt.2005.073890](https://doi.org/10.1136/hrt.2005.073890)
11. Herscovici R, Sedlak T, Wei J, Pepine CJ, Handberg E, Merz CNB. Ischemia and no obstructive coronary artery disease (INOCA): what is the risk? *J Am Heart Assoc*. 2018;7:e008868. doi: [10.1161/JAHA.118.008868](https://doi.org/10.1161/JAHA.118.008868)
12. Jespersen L, Hvelplund A, Abildstrøm SZ, Pedersen F, Galatius S, Madsen JK, Jørgensen E, Kelbæk H, Prescott E. Stable angina pectoris with no obstructive coronary artery disease is associated with increased risks of major adverse cardiovascular events. *Eur Heart J*. 2011;33:734–744. doi: [10.1093/eurheartj/ehr331](https://doi.org/10.1093/eurheartj/ehr331)
13. Cardiovascular stress test cost and procedure information. New Choice Health. 2023. Accessed February 17, 2023. <https://www.newchoicehealth.com/>
14. Jassawalla F. Cardiac PET scans: who needs them and why? CURA4U. 2021. Accessed September 3, 2021. <https://cura4u.com/blog/cardiac-pet-scans-who-needs-them-and-why>
15. Callison K. Medicare managed care spillovers and treatment intensity. *Health Econ*. 2016;25:873–887. doi: [10.1002/hec.3191](https://doi.org/10.1002/hec.3191)
16. Siminoff LA, Hausmann LR, Ibrahim S. Barriers to obtaining diagnostic testing for coronary artery disease among veterans. *Am J Public Health*. 2008;98:2207–2213. doi: [10.2105/AJPH.2007.123224](https://doi.org/10.2105/AJPH.2007.123224)
17. Hou K, Burch KS, Majumdar A, Shi H, Mancuso N, Wu Y, Sankararaman S, Pasaniuc B. Accurate estimation of SNP-heritability from biobank-scale data irrespective of genetic architecture. *Nat Genet*. 2019;51:1244–1251. doi: [10.1038/s41588-019-0465-0](https://doi.org/10.1038/s41588-019-0465-0)
18. The Blueprint consortium, Bock C, Halbritter F, Carmona FJ, Tierling S, Datlinger P, Assenov Y, Berdasco M, Bergmann AK, Booher K, et al. Quantitative comparison of DNA methylation assays for biomarker development and clinical applications. *Nat Biotechnol*. 2016;34:726–737. doi: [10.1038/nbt.3605](https://doi.org/10.1038/nbt.3605)
19. Tsao CW, Vasan RS. Cohort profile: the Framingham Heart Study (FHS): overview of milestones in cardiovascular epidemiology. *Int J Epidemiol*. 2015;44:1800–1813. doi: [10.1093/ije/dyv337](https://doi.org/10.1093/ije/dyv337)
20. Dawber TR, Meadors GF, Moore FE Jr. Epidemiological approaches to heart disease: the Framingham study. *Am J Public Health*. 1951;41:279–286. doi: [10.2105/AJPH.41.3.279](https://doi.org/10.2105/AJPH.41.3.279)
21. Cupples L, D'Agostino R, Kiely D. The Framingham Heart Study, section 35. An epidemiological investigation of cardiovascular disease survival following cardiovascular events: 30 year follow-up. Lung and Blood Institute; 1988;35:1–454.
22. Muhlestein JB, May HT, Bair TL, Prescott MF, Horne BD, White R, Anderson JL. Relation of elevated plasma renin activity at baseline to cardiac events in patients with angiographically proven coronary artery disease. *Am J Cardiol*. 2010;106:764–769. doi: [10.1016/j.amjcard.2010.04.040](https://doi.org/10.1016/j.amjcard.2010.04.040)
23. Taylor GS, Muhlestein JB, Wagner GS, Bair TL, Li P, Anderson JL. Implementation of a computerized cardiovascular information system in a private hospital setting. *Am Heart J*. 1998;136:792–803. doi: [10.1016/S0002-8703\(98\)70123-1](https://doi.org/10.1016/S0002-8703(98)70123-1)
24. Hamilton CM, Strader LC, Pratt JG, Maiese D, Hendershot T, Kwok RK, Hammond JA, Huggins W, Jackman D, Pan H, et al. The PhenX Toolkit: get the most from your measures. *Am J Epidemiol*. 2011;174:253–260. doi: [10.1093/aje/kwr193](https://doi.org/10.1093/aje/kwr193)
25. Philibert R, Dogan M, Beach SRH, Mills JA, Long JD. AHRR methylation predicts smoking status and smoking intensity in both saliva and blood DNA. *Am J Genet*. 2019;183:51–60. doi: [10.1002/ajmg.b.32760](https://doi.org/10.1002/ajmg.b.32760)
26. Pidsley R, Y, Wong CC, Volta M, Lunnon K, Mill J, Schalkwyk LC. A data-driven approach to preprocessing Illumina 450k methylation array data. *BMC Genomics*. 2013;14:1–10. doi: [10.1186/1471-2164-14-293](https://doi.org/10.1186/1471-2164-14-293)
27. Triche Jr T. *FDB.InfiniumMethylation.hg19: Annotation package for Illumina Infinium DNA methylation probes*. R package version 2.2.0. 2014. Accessed July 1, 2022. <https://bioconductor.org/packages/release/data/annotation/html/FDB.InfiniumMethylation.hg19.html>
28. Davis S, Du P, Bilke S, Triche JT, Bootwalla M. *Methylumi: Handle Illumina methylation data*. R package version 2.22.0. 2017. Accessed July 1, 2022. <https://bioconductor.statistik.tu-dortmund.de/packages/3.5/bioc/html/methylumi.html>
29. Dogan M, Beach S, Simons R, Lendasse A, Penaluna B, Philibert R. Blood-based biomarkers for predicting the risk for five-year incident coronary heart disease in the Framingham Heart Study via machine learning. *Genes*. 2018;9:641. doi: [10.3390/genes9120641](https://doi.org/10.3390/genes9120641)
30. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007;81:559–575. doi: [10.1086/519795](https://doi.org/10.1086/519795)
31. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V. Scikit-learn: machine learning in python. *J Mach Learn Res*. 2011;12:2825–2830.
32. Dalcin LD, Paz RR, Kler PA, Cosimo A. Parallel distributed computing using Python. *Adv Water Resour*. 2011;34:1124–1139. doi: [10.1016/j.advwatres.2011.04.013](https://doi.org/10.1016/j.advwatres.2011.04.013)
33. Breiman L. Pasting small votes for classification in large databases and on-line. *Mach Learn*. 1999;36:85–103. doi: [10.1023/A:1007563306331](https://doi.org/10.1023/A:1007563306331)
34. Breiman L. Bagging predictors. *Mach Learn*. 1996;24:123–140. doi: [10.1007/BF00058655](https://doi.org/10.1007/BF00058655)
35. Louppe G, Geurts P. Ensembles on random patches. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. Springer; 2012:346–361.
36. Chen C, Liaw A, Breiman L. *Using random forest to learn imbalanced data*. 2004. <https://statistics.berkeley.edu/sites/default/files/tech-reports/666.pdf>
37. Opitz DW, Maclin RF. An empirical evaluation of bagging and boosting for artificial neural networks. In: *Proceedings of International Conference on Neural Networks (ICNN'97)* (Vol. 3). New York, NY: IEEE; 1997: 1401–1405.
38. Hido S, Kashima H, Takahashi Y. Roughly balanced bagging for imbalanced data. *Stat Anal Data Min*. 2009;2:412–426. doi: [10.1002/sam.10061](https://doi.org/10.1002/sam.10061)
39. Wang S, Yao X. Diversity analysis on imbalanced data sets by using ensemble models. *IEEE Int Symp Comput Intell Inform*. 2009;2009: 324–331.
40. Lemaître G, Nogueira F, Aridas CK. Imbalanced-learn: a Python toolbox to tackle the curse of imbalanced datasets in machine learning. *J Mach Learn Res*. 2017;18:559–563.
41. Philibert R, Miller S, Noel A, Dawes K, Papworth E, Black DW, Beach SRH, Long JD, Mills JA, Dogan M. A four marker digital PCR toolkit for detecting heavy alcohol consumption and the effectiveness of its treatment. *J Insur Med*. 2019;48:90–102. doi: [10.17849/insm-48-1-1-1](https://doi.org/10.17849/insm-48-1-1-1)
42. Philibert R, Dogan M, Noel A, Miller S, Krukow B, Papworth E, Cowley J, Long JD, Beach SR, Black DW. Dose response and prediction

- characteristics of a methylation sensitive digital PCR assay for cigarette consumption in adults. *Front Genet.* 2018;9:137. doi: 10.3389/fgene.2018.00137
43. Muret K, Désert C, Lagoutte L, Boutin M, Gondret F, Zerjal T, Lagarrigue S. Long noncoding RNAs in lipid metabolism: literature review and conservation analysis across species. *BMC Genomics.* 2019;20:882. doi: 10.1186/s12864-019-6093-3
 44. Gil N, Ulitsky I. Regulation of gene expression by cis-acting long non-coding RNAs. *Nat Rev Genet.* 2020;21:102–117. doi: 10.1038/s41576-019-0184-5
 45. Waterham HR, Koster J, Romeijn GJ, Hennekam RCM, Vreken P, Andersson HC, FitzPatrick DR, Kelley RI, Wanders RJA. Mutations in the 3 β -hydroxysterol δ 24-reductase gene cause desmosterolosis, an autosomal recessive disorder of cholesterol biosynthesis. *Am J Hum Genet.* 2001;69:685–694. doi: 10.1086/323473
 46. Lambert G, Charlton F, Rye K-A, Piper DE. Molecular basis of PCSK9 function. *Atherosclerosis.* 2009;203:1–7. doi: 10.1016/j.atherosclerosis.2008.06.010
 47. Daugherty A, Dunn JL, Rateri DL, Heinecke JW. Myeloperoxidase, a catalyst for lipoprotein oxidation, is expressed in human atherosclerotic lesions. *J Clin Invest.* 1994;94:437–444. doi: 10.1172/JCI117342
 48. Fernández-Sanlés A, Sayols-Baixeras S, Subirana I, Sentí M, Pérez-Fernández S, de Castro MM, Esteller M, Marrugat J, Elosua R. DNA methylation biomarkers of myocardial infarction and cardiovascular disease. *Clin Epigenetics.* 2021;13:86. doi: 10.1186/s13148-021-01078-6
 49. Dawes K, Andersen A, Reimer R, Mills JA, Hoffman E, Long JD, Miller S, Philibert R. The relationship of smoking to cg05575921 methylation in blood and saliva DNA samples from several studies. *Sci Rep.* 2021;11:21627. doi: 10.1038/s41598-021-01088-7
 50. Philibert RA, Beach SR, Lei M-K, Brody GH. Changes in DNA methylation at the aryl hydrocarbon receptor repressor may be a new biomarker for smoking. *Clin Epigenetics.* 2013;5:1–8. doi: 10.1186/1868-7083-5-19
 51. Fang F, Andersen AM, Philibert R, Hancock DB. Epigenetic biomarkers for smoking cessation. *Add Neurosci.* 2023;6:100079. doi: 10.1016/j.addicn.2023.100079
 52. Vincent C, Tarbouriech N, Härtlein M. Genomic organization, cDNA sequence, bacterial expression, and purification of human seryl-tRNA synthase. *Europ J Mol Biol Biochem.* 1997;250:77–84. doi: 10.1111/j.1432-1033.1997.00077.x
 53. Herzog W, Müller K, Huiskens J, Stainier DY. Genetic evidence for a non-canonical function of seryl-tRNA synthetase in vascular development. *Circ Res.* 2009;104:1260–1266. doi: 10.1161/CIRCRESAHA.108.191718
 54. Zheng Y, Joyce BT, Hwang S-J, Ma J, Liu L, Allen NB, Krefman AE, Wang J, Gao T, Nannini DR, et al. Association of cardiovascular health through young adulthood with genome-wide DNA methylation patterns in midlife: the CARDIA study. *Circulation.* 2022;146:94–109. doi: 10.1161/CIRCULATIONAHA.121.055484
 55. Mendelson MM, Marioni RE, Joehanes R, Liu C, Hedman ÅK, Aslibekyan S, Demerath EW, Guan W, Zhi D, Yao C, et al. Association of body mass index with DNA methylation and gene expression in blood cells and relations to cardiometabolic disease: a Mendelian randomization approach. *PLoS Med.* 2017;14:e1002215. doi: 10.1371/journal.pmed.1002215
 56. Wahl S, Drong A, Lehne B, Loh M, Scott WR, Kunze S, Tsai P-C, Ried JS, Zhang W, Yang Y. Epigenome-wide association study of body mass index, and the adverse outcomes of adiposity. *Nature.* 2017;541:81–86. doi: 10.1038/nature20784
 57. Handzik MK, Gengatharan JM, Frizzi KE, McGregor GH, Martino C, Rahman G, Gonzalez A, Moreno AM, Green CR, Guernsey LS, et al. Insulin-regulated serine and lipid metabolism drive peripheral neuropathy. *Nature.* 2023;614:118–124. doi: 10.1038/s41586-022-05637-6
 58. Hornemann T. Serine deficiency causes complications in diabetes. *Nature.* 2023;614:42–43. doi: 10.1038/d41586-023-00054-9
 59. Jiang J, Li B, He W, Huang C. Dietary serine supplementation: friend or foe? *Curr Opin Pharmacol.* 2021;61:12–20. doi: 10.1016/j.coph.2021.08.011
 60. Naslavsky N, Caplan S. EHD proteins: key conductors of endocytic transport. *Trends Cell Biol.* 2011;21:122–131. doi: 10.1016/j.tcb.2010.10.003
 61. Meeks KAC, Henneman P, Venema A, Addo J, Bahendeka S, Burr T, Danquah I, Galbete C, Mannens MMAM, Mockenhaupt FP, et al. Epigenome-wide association study in whole blood on type 2 diabetes among sub-Saharan african individuals: findings from the RODAM study. *Int J Epidemiol.* 2018;48:58–70. doi: 10.1093/ije/dyy171
 62. Wang Z, Gao W, Wang B, Cao W, Lv J, Yu C, Pang Z, Cong L, Wang H, Wu X. Correlation between fasting plasma glucose, HbA1c and DNA methylation in adult twins. Article in Chinese. *Beijing Da Xue Xue Bao Yi Xue Ban.* 2020;52:425–431. doi: 10.19723/j.issn.1671-167X.2020.03.005
 63. Kurian SM, Ferreri K, Wang C-H, Todorov I, Al-Abdullah IH, Rawson J, Mullen Y, Salomon DR, Kandeel F. Gene expression signature predicts human islet integrity and transplant functionality in diabetic mice. *PLoS One.* 2017;12:e0185331. doi: 10.1371/journal.pone.0185331
 64. Gomes KP, Jadli AS, de Almeida LGN, Ballasy NN, Edalat P, Shandilya R, Young D, Belke D, Shearer J, Dufour A, et al. Proteomic analysis suggests altered mitochondrial metabolic profile associated with diabetic cardiomyopathy. *Front Cardiovasc Med.* 2022;9:791700. doi: 10.3389/fcvm.2022.791700
 65. Korbecki J, Barczak K, Gutowska I, Chlubek D, Baranowska-Bosiacka I. CXCL1: gene, promoter, regulation of expression, mRNA stability, regulation of activity in the intercellular space. *Int J Mol Sci.* 2022;23:792. doi: 10.3390/ijms23020792
 66. Zhao W, Siegel D, Biton A, Tonqueze OL, Zaitlen N, Ahituv N, Erle DJ. Crispr-Cas9-mediated functional dissection of 3'-UTRs. *Nucleic Acids Res.* 2017;45:10800–10810. doi: 10.1093/nar/gkx675
 67. Navarro E, Mallén A, Hueso M. Dynamic variations of 3'UTR length reprogram the mRNA regulatory landscape. *Biomedicine.* 2021;9:1560. doi: 10.3390/biomedicines9111560
 68. Kata D, Földesi I, Feher LZ, Hackler L, Puskas LG, Gulya K. A novel pleiotropic effect of aspirin: beneficial regulation of pro- and anti-inflammatory mechanisms in microglial cells. *Brain Res Bull.* 2017;132:61–74. doi: 10.1016/j.brainresbull.2017.05.009
 69. Ittaman SV, VanWormer JJ, Rezakalla SH. The role of aspirin in the prevention of cardiovascular disease. *Clin Med Res.* 2014;12:147–154. doi: 10.3121/cmr.2013.1197
 70. Raju NC, Eikelboom JW. The aspirin controversy in primary prevention. *Curr Opin Cardiol.* 2012;27:499–507. doi: 10.1097/HCO.0b013e328356ae95
 71. Kalnina I, Zaharenko L, Vaivade I, Rovite V, Nikitina-Zake L, Peculis R, Fridmanis D, Geldnere K, Jacobsson JA, Almen MS, et al. Polymorphisms in FTO and near TMEM18 associate with type 2 diabetes and predispose to younger age at diagnosis of diabetes. *Gene.* 2013;527:462–468. doi: 10.1016/j.gene.2013.06.079
 72. Zhao W, Rasheed A, Tikkanen E, Lee J-J, Butterworth AS, Howson JMM, Assimes TL, Chowdhury R, Orho-Melander M, Damrauer S, et al. Identification of new susceptibility loci for type 2 diabetes and shared etiological pathways with coronary heart disease. *Nat Genet.* 2017;49:1450–1457. doi: 10.1038/ng.3943
 73. Lenselink C, Ties D, Pleijhuis R, van der Harst P. Validation and comparison of 28 risk prediction models for coronary artery disease. *Eur J Prev Cardiol.* 2022;29:666–674. doi: 10.1093/eurjpc/zwab095
 74. Zhang X, Wang C, He D, Cheng Y, Yu L, Qi D, Li B, Zheng F. Identification of DNA methylation-regulated genes as potential biomarkers for coronary heart disease via machine learning in the Framingham Heart Study. *Clin Epigenetics.* 2022;14:122. doi: 10.1186/s13148-022-01343-2
 75. Kunadian V, Chieffo A, Camici PG, Berry C, Escaned J, Maas AHM, Prescott E, Karam N, Appelman Y, Fraccaro C, et al. An EAPCI expert consensus document on Ischaemia with non-obstructive coronary arteries in collaboration with European Society of Cardiology Working Group on Coronary Pathophysiology & Microcirculation. Endorsed by Coronary Vasomotor Disorders International Study Group. *Eur Heart J.* 2020;41:3504–3520. doi: 10.1093/eurheartj/ehaa503
 76. Luu JM, Malhotra P, Cook-Wiens G, Pepine CJ, Handberg EM, Reis SE, Reichel N, Bittner V, Wei J, Kelsey SF. Long-term adverse outcomes in Black women with ischemia and no obstructive coronary artery disease: a study of the WISE (Women's Ischemia Syndrome Evaluation) cohort. *Circulation.* 2023;147:617–619. doi: 10.1161/CIRCULATIONAHA.122.063466
 77. Ford TJ, Corcoran D, Berry C. Stable coronary syndromes: pathophysiology, diagnostic advances and therapeutic need. *Heart.* 2018;104:284–292. doi: 10.1136/heartjnl-2017-311446
 78. Ford TJ, Berry C. How to diagnose and manage angina without obstructive coronary artery disease: lessons from the British Heart Foundation CorMicA trial. *Intervent Cardiol.* 2019;14:76–82. doi: 10.15420/icr.2019.04.R1

79. James MT, Samuel SM, Manning MA, Tonelli M, Ghali WA, Faris P, Knudtson ML, Pannu N, Hemmelgarn BR. Contrast-induced acute kidney injury and risk of adverse clinical outcomes after coronary angiography. *Circ Cardiovasc Interv.* 2013;6:37–43. doi: [10.1161/CIRCINTERVENTIONS.112.974493](https://doi.org/10.1161/CIRCINTERVENTIONS.112.974493)
80. Clinical implications of revised pooled cohort equations for estimating atherosclerotic cardiovascular disease risk. *Ann Intern Med.* 2018;169:20–29. doi: [10.7326/M17-3011](https://doi.org/10.7326/M17-3011)
81. Ridker PM, Cook NR. The pooled cohort equations 3 years on. *Circulation.* 2016;134:1789–1791. doi: [10.1161/CIRCULATIONAHA.116.024246](https://doi.org/10.1161/CIRCULATIONAHA.116.024246)
82. Dedeurwaerder S, Defrance M, Bizet M, Calonne E, Bontempi G, Fuks F. A comprehensive overview of Infinium HumanMethylation450 data processing. *Brief Bioinform.* 2013;15:929–941. doi: [10.1093/bib/bbt054](https://doi.org/10.1093/bib/bbt054)
83. Cheung K, Burgers MJ, Young DA, Cockell S, Reynard LN. Correlation of Infinium HumanMethylation450k and MethylationEPIC BeadChip arrays in cartilage. *Epigenetics.* 2020;15:594–603. doi: [10.1080/15592294.2019.1700003](https://doi.org/10.1080/15592294.2019.1700003)
84. Noble AJ, Pearson JF, Boden JM, Horwood LJ, Gemmell NJ, Kennedy MA, Osborne AJ. A validation of Illumina EPIC array system with bisulfite-based amplicon sequencing. *PeerJ.* 2021;9:e10762. doi: [10.7717/peerj.10762](https://doi.org/10.7717/peerj.10762)
85. Kruppa J, Sieg M, Richter G, Pohrt A. Estimands in epigenome-wide association studies. *Clin Epigenetics.* 2021;13:98. doi: [10.1186/s13148-021-01083-9](https://doi.org/10.1186/s13148-021-01083-9)
86. Philibert R, Dawes K, Moody J, Hoffman R, Sieren J, Long J. Using cg05575921 methylation to predict lung cancer risk: a potentially bias-free precision epigenetics approach. *Epigenetics.* 2022;17:2096–2108. doi: [10.1080/15592294.2022.2108082](https://doi.org/10.1080/15592294.2022.2108082)
87. Philibert R, Dogan M, Noel A, Miller S, Krukow B, Papworth E, Cowley J, Knudsen A, Beach SRH, Black D. Genome-wide and digital polymerase chain reaction epigenetic assessments of alcohol consumption. *Am J Med Genet B Neuropsychiatr Genet.* 2018;177:479–488. doi: [10.1002/ajmg.b.32636](https://doi.org/10.1002/ajmg.b.32636)