two-year, or ten-year) as well as the incidence or prevalence of other types of CVD including, without limitation, CHD, stroke, arrhythmia, cardiac arrest, and congestive heart failure. Specifically, methods and compositions for determining the methylation status of at least one CpG locus and at least one single nucleotide polymorphism (SNP) are described.

In one example, kits for determining methylation status of at least one CpG dinucleotide and a genotype of at least one single-nucleotide polymorphism (SNP) are provided. Such kits typically include at least one first nucleic acid primer at least 8 nucleotides in length that is complementary to a bisulfite-converted nucleic acid sequence comprising a first CpG dinucleotide at a GC locus selected from the group consisting of cg00300879, cg09552548, and cg14789911 or at a second CpG dinucleotide in linkage disequilibrium with the first CpG dinucleotide at a GC locus selected from the group consisting of cg00300879, cg09552548, and cg14789911, wherein the linkage disequilibrium has a value of $R>0.3$, wherein the at least one first nucleic acid primer detects a methylated or unmethylated CpG dinucleotide, and at least one second nucleic acid primer at least 8 nucleotides in length that is complementary to a DNA sequence or a bisulfite-converted DNA sequence of a first SNP selected from the group consisting of rs11716050, rs6560711, rs3735222, rs6820447, and rs9638144 or a second SNP in linkage disequilibrium with the first SNP selected from the group consisting of rs11716050, rs6560711, rs3735222, rs6820447, and rs9638144, wherein the linkage disequilibrium has a value of $R>0.3$.

In some embodiments, the at least one first nucleic acid primer detects the unmethylated CpG dinucleotide. In some examples, the at least one first nucleic acid primer detects the methylated CpG dinucleotide.

In some examples, the kits described herein further including at least a third nucleic acid primer at least 8 nucleotides in length that is complementary to a nucleic acid sequence upstream of the CpG dinucleotide. In some examples, the kits further include at least a third nucleic acid primer at least 8 nucleotides in length that is complementary to a nucleic acid sequence downstream of the CpG dinucleotide.

In some examples, the at least one first nucleic acid primer comprises one or more nucleotide analogs. In some examples, the at least one first nucleic acid primer comprises one or more synthetic or non-natural nucleotides.

In some examples, the kits described herein further include a solid substrate to which the at least one first nucleic acid primer is bound. In some examples, the substrate is a polymer, glass, semiconductor, paper, metal, gel or hydrogel. In some examples, the solid substrate is a microarray or microfluidics card.

In some examples, the kits described herein further include a detectable label.

In another example, methods of determining the presence of biomarkers associated with predicting CHD in a biological sample from a patient is provided. Such methods typically include (a) providing a first portion of the biological sample and a second portion of the biological sample, wherein the nucleic acid from at least the first portion is bisulfite converted; (b) contacting the first portion of the biological sample with a first oligonucleotide primer at least 8 nucleotides in length that is complementary to a sequence that comprises a first CpG dinucleotide at a GC locus selected from the group consisting of cg00300879, cg09552548, and cg14789911, or a second CpG dinucleotide in linkage disequilibrium with the first CpG dinucleotide at a GC locus selected from the group consisting of cg00300879, cg09552548, and cg14789911, wherein the linkage disequilibrium has a value of R>0.3, wherein the first nucleic acid primer detects a methylated or unmethylated CpG dinucleotide; and (c) contacting the second portion of the biological sample with a nucleic acid primer at least 8 nucleotides in length that is complementary to a DNA sequence or a bisulfite-converted DNA sequence of a first SNP selected from the group consisting of rs11716050, rs6560711, rs3735222, rs6820447, and rs9638144 or a second SNP in linkage disequilibrium with the first SNP selected from the group consisting of rs11716050, rs6560711, rs3735222, rs6820447, and rs9638144, wherein the linkage disequilibrium has a value of R>0.3. Generally, the percentage of methylation of the CpG dinucleotide at the GC locus selected from the group consisting of cg00300879, cg09552548, and cg14789911, and the identity of the nucleotide at the first SNP selected from the group consisting of rs11716050, rs6560711, rs3735222, rs6820447, and rs9638144 or the second SNP in linkage disequilibrium with the first SNP are biomarkers associated with the incidence of CHD.

In some examples, the biological sample is selected from the group consisting of blood and saliva.

In some examples, the at least one first nucleic acid primer detects the unmethylated CpG dinucleotide. In some examples, the at least one first nucleic acid primer detects the methylated CpG dinucleotide.

In some examples, the at least one first nucleic acid primer comprises one or more nucleotide analogs. In some examples, the at least one first nucleic acid primer comprises one or more synthetic or non-natural nucleotides.

In some examples, the window of incidence is three years.

In still a further example, methods of determining the presence of a biomarker associated with CHD in a patient sample are provided. Such methods typically include (a) isolating nucleic acid sample from the patient sample, (b) performing a genotyping assay on a first portion of the nucleic acid sample to detect the presence of at least one SNP, wherein the at least one SNP is a first SNP from Appendix C and/or is a second SNP in linkage disequilibrium (R>0.3) with a first SNP from Appendix C to obtain genotype data; and/or (c) bisulfite converting the nucleic acid in a second portion of the nucleic acid and performing methylation assessment on the second portion of the nucleic acid sample to detect methylation status of at least one CpG site from Appendix A and/or a CpG site collinear (R>0.3) with a CpG from Appendix A to obtain methylation data; and (d) inputing the genotype data from step (b) and/or methylation data from step (c) into an algorithm that accounts for at least one SNP main effect and/or at least one CpG main effect and/or at least one interaction effect, wherein the algorithm is a machine learning algorithm capable of accounting for linear and non-linear effects.

In some examples, the at least one interaction effect is selected from the group consisting of a gene-environment interaction (SNPxCpG) effect, a gene-gene interaction (SNPxSNP) effect, and an environment-environment interaction (CpGxCpG) effect. In some examples, the at least one interaction effect is a gene-environment interaction effect (SNPxCpG) between a CpG site from Appendix A or a CpG site that is collinear (R>0.3) with a CpG site from Appendix A and a SNP from Appendix C or a SNP within moderate linkage disequilibrium (R>0.3) from a SNP from Appendix C. In some examples, the at least one interaction effect is an environment-environment interaction effect (CpGxCpG) between at least two CpG sites from Appendix A.

In some examples, one or both of the at least two CpG sites are collinear (R>0.3) with one or both of the at least two CpG sites from Appendix A. In some examples, the at least one interaction effect is a gene-gene interaction effect (SNPxSNP) between at least two SNPs from Appendix C. In some examples, one or both of the at least two SNPs are collinear (R>0.3) with one or both of the at least two SNPs from Appendix C.

In some examples, the biological sample is a saliva sample.

In another example, systems for determining methylation status of at least one CpG dinucleotide and a genotype of at least one single-nucleotide polymorphism (SNP) are provided. Such systems typically include: a nucleic acid isolation module configured to isolate a nucleic acid sample from a subject sample; a genotyping assay module configured to perform a genotyping assay on a first portion of the nucleic acid sample to detect the presence of at least one SNP, wherein the at least one SNP is a first SNP from Appendix C and/or is a second SNP in linkage disequilibrium (R>0.3) with a first SNP from Appendix C to obtain genotype data; a methylation assay module configured to bisulfite convert the nucleic acid in a second portion of the nucleic acid and perform a methylation assessment on a second portion of the nucleic acid sample to detect methylation status of at least one CpG site from Appendix A and/or a CpG site collinear (R>0.3) with a CpG from Appendix A to obtain methylation data; and an identification system configured to account for at least one SNP main effect and/or at least one CpG main effect and/or at least one interaction effect based on the genotype data from step (b) and/or methylation data from step (c).

In some examples, such systems further include an output module configured to provide an output based on an identification by the identification system, wherein the identification accounts for at least one SNP main effect and/or at least one CpG main effect and/or at least one interaction effect based on the genotype data from step (b) and/or methylation data from step (c).

In some examples, the algorithm is a machine learning algorithm capable of accounting for linear and non-linear effects.

In yet another example, non-transitory computer-readable media storing instructions executable by a processing device to perform operations are provided. Such operations typically include accounting for at least one SNP main effect and/or at least one CpG main effect and/or at least one interaction effect based on genotype data and/or methylation data,

wherein: (i) the genotype data is based on a genotyping assay on a first portion of a nucleic acid sample isolated from a subject sample to detect the presence of at least one SNP, wherein the at least one SNP is a first SNP from Appendix C and/or is a second SNP in linkage disequilibrium (R>0.3) with a first SNP from Appendix C to obtain the genotype data; and (ii) the methylation data is based on a methylation assay on a bisulfite converted nucleic acid in a second portion of the nucleic acid sample to detect methylation status of at least one CpG site from Appendix A and/or a CpG site collinear (R>0.3) with a CpG from Appendix A to obtain methylation data.

In some examples, the operations further include providing an output based on the accounting. Representative outputs, without limitation, include one or more of storing a report based on the accounting to another non-transitory computer-readable medium, modifying a display based on the accounting, triggering an audible alert based on the accounting, triggering a haptic or vibratory alert based on the accounting, triggering the printing of a report based on the accounting, or triggering the delivery of a therapeutic based on the accounting.

In one aspect, the present disclosure provides a kit for determining methylation status of CpG dinucleotides and genotype of single-nucleotide polymorphisms (SNPs), the kit comprising:

at least three first nucleic acid primers that are each at least 8 nucleotides in length and complementary to a bisulfite-converted nucleic acid sequence comprising CpG dinucleotides at GC loci cg00300879, cg09552548, and cg14789911, wherein each of the at least three first nucleic acid primers detect a methylated or unmethylated CpG dinucleotide, and

at least five second nucleic acid primers that are each at least 8 nucleotides in length and complementary to a DNA sequence or a bisulfite-converted DNA sequence of SNPs rs11716050, rs6560711, rs3735222, rs6820447, and rs9638144.

In another aspect, the present disclosure provides a method of determining the presence of biomarkers associated with predicting a three-year incidence of CVD using a biological sample from a subject, comprising:

(a) providing a first portion of the biological sample and a second portion of the biological sample, wherein the nucleic acid from at least the first portion is bisulfite converted;

(b) contacting the first portion of the biological sample with at least three first oligonucleotide primers that are each at least 8 nucleotides in length and complementary to sequences that comprise CpG dinucleotides at GC loci cg00300879, cg09552548, and cg14789911, wherein each of the at least three first nucleic acid primers detect a methylated or unmethylated CpG dinucleotide; and

(c) contacting the second portion of the biological sample with at least five second nucleic acid primers that are each at least 8 nucleotides in length and complementary to a DNA sequence or a bisulfite-converted DNA sequence of SNPs rs11716050, rs6560711, rs3735222, rs6820447, and rs9638144,

wherein the percentage of methylation of the CpG dinucleotides at the GC loci cg00300879, cg09552548, and cg14789911, and the identity of the nucleotide at the SNPs rs11716050, rs6560711, rs3735222, rs6820447, and rs9638144 are biomarkers associated with the three-year incidence of CVD.

The integrated genetic-epigenetic model described herein provides several advantages and benefits. The first is the overall sensitivity across cohorts. On average, in the Intermountain Healthcare (IM) cohort, the typical risk calculators accurately identify 5 out of 10 individuals at high risk for an incident event compared to the integrated genetic-epigenetic model described herein, which accurately identifies 7 out of 10 individuals. The second is with respect to the performance of standard risk calculators by gender. On average, in the IM cohort, the typical risk calculators accurately identify 5 of 10 men and 4 of 10 women at risk for an incident event. On average, in the IM cohort, the integrated genetic-epigenetic tool described herein accurately identifies 7 of 10 men and 7 of 10 women at risk for an incident event. Thus, the integrated genetic-epigenetic model described herein does not exhibit gender gap in its ability to identify men and women at risk for an incident event.

Unless otherwise defined, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which the methods and compositions of matter belong. Although methods and materials similar or equivalent to those described herein can be used in the practice or testing of the methods and compositions

WHAT IS CLAIMED IS:

1. A kit for determining methylation status of CpG dinucleotides and genotype of single-nucleotide polymorphisms (SNPs), the kit comprising:

at least three first nucleic acid primers that are each at least 8 nucleotides in length and complementary to a bisulfite-converted nucleic acid sequence comprising CpG dinucleotides at GC loci cg00300879, cg09552548, and cg14789911, wherein each of the at least three first nucleic acid primers detect a methylated or unmethylated CpG dinucleotide, and

at least five second nucleic acid primers that are each at least 8 nucleotides in length and complementary to a DNA sequence or a bisulfite-converted DNA sequence of SNPs rs11716050, rs6560711, rs3735222, rs6820447, and rs9638144.

2. The kit of claim 1, wherein at least one of the first nucleic acid primers detect the unmethylated CpG dinucleotide.

3. The kit of claim 1, wherein at least one of the first nucleic acid primers detect the methylated CpG dinucleotide.

4. The kit of any of claims 1-3, wherein at least one of the first nucleic acid primers comprise one or more nucleotide analogs.

5. The kit of any of claims 1-4, wherein at least one of the first nucleic acid primers comprise one or more synthetic or non-natural nucleotides.

6. The kit of any one of claims 1-5, further comprising a solid substrate to which at least one of the first nucleic acid primers is bound.

7. The kit of claim 6, wherein the substrate is a polymer, glass, semiconductor, paper, metal, gel or hydrogel.

8. The kit of claim 6, wherein the solid substrate is a microarray or microfluidics card.

9. The kit of any one of claims 1-8, further comprising a detectable label.

10. The kit of any one of claims 1-9, further comprising at least a third nucleic acid primer at least 8 nucleotides in length that is complementary to a nucleic acid sequence upstream of the CpG dinucleotide.

11. The kit of any one of claims 1-9, further comprising at least a third nucleic acid primer at least 8 nucleotides in length that is complementary to a nucleic acid sequence downstream of the CpG dinucleotide.

12. A method of determining the presence of biomarkers associated with predicting a three-year incidence of CVD using a biological sample from a subject, comprising:

(a) providing a first portion of the biological sample and a second portion of the biological sample, wherein the nucleic acid from at least the first portion is bisulfite converted;

(b) contacting the first portion of the biological sample with at least three first oligonucleotide primers that are each at least 8 nucleotides in length and complementary to sequences that comprise CpG dinucleotides at GC loci cg00300879, cg09552548, and cg14789911, wherein each of the at least three first nucleic acid primers detect a methylated or unmethylated CpG dinucleotide; and

(c) contacting the second portion of the biological sample with at least five second nucleic acid primers that are each at least 8 nucleotides in length and complementary to a DNA sequence or a bisulfite-converted DNA sequence of SNPs rs11716050, rs6560711, rs3735222, rs6820447, and rs9638144,

wherein the percentage of methylation of the CpG dinucleotides at the GC loci cg00300879, cg09552548, and cg14789911, and the identity of the nucleotide at the SNPs

rs11716050, rs6560711, rs3735222, rs6820447, and rs9638144 are biomarkers associated with the three-year incidence of CVD.

13. The method of claim 12, wherein the biological sample is selected from the group consisting of blood and saliva.

14. The method of claim 12, wherein at least one of the first nucleic acid primers detect the unmethylated CpG dinucleotide.

15. The method of claim 12, wherein at least one of the first nucleic acid primers detect the methylated CpG dinucleotide.

16. The method of claim 12, wherein at least one of the first nucleic acid primers comprise one or more nucleotide analogs.

17. The method of claim 12, wherein at least one of the first nucleic acid primers comprise one or more synthetic or non-natural nucleotides.

18. The method of claim 12, wherein the window of incidence is three years.