

METHODS AND COMPOSITIONS FOR PREDICTING AND/OR MONITORING CARDIOVASCULAR DISEASE AND INTERVENTIONS THEREFOR

TECHNICAL FIELD

5 This disclosure generally relates to methods and compositions related to predicting cardiovascular disease (CVD) in an individual.

BACKGROUND

Cardiovascular disease (CVD), and particularly coronary heart disease (CHD), is the
10 most common type of heart disease and was responsible for over 360,000 deaths in the United States in 2017. In order to decrease this toll, a number of risk estimators and detection methods have been developed to better identify those with or at risk for CVD, including CHD. Beginning with the Framingham Risk Score (FRS) and more recently, the ASCVD Pooled Cohort Equation (PCE), these tools capture variance in key physiological parameters,
15 such as serum lipid levels, known to be associated with risk for CVD, including CHD. Similarly, for detection, methods such as stress echocardiogram and Coronary Computed Tomography Angiography (CCTA) are used.

Despite the magnitude of these efforts, current risk estimators and detection tests often lack in sensitivity and specificity, and often are not as accessible due to cost and the
20 need to schedule an in-person clinical visit that may take weeks. Furthermore, some of the current risk estimators and detection methods such as catheterization may have severe side effects such as stroke and heart attack. As a result, there is a need for alternative stratification, detection and management approaches for CVD that have minimal risks, are scalable, and provide actionable insights.

SUMMARY

25 Methods and compositions for predicting the presence and/or severity (e.g., level of obstruction) of cardiovascular disease (CVD) are provided, and methods and compositions

for managing, monitoring, and/or treating CVD are provided. For example, methods and compositions for predicting coronary heart disease (CHD) are described herein. The general principals apply to windows of incidence (e.g., one-month, six-month, two-year, or ten-year) as well as the incidence, prevalence, or severity of other types of CVD including, without
5 limitation, CHD, stroke, arrhythmia, cardiac arrest, and congestive heart failure. The same general principals also apply to survival of CVD or CVD events as well as to the management of CVD or CVD events, including but not limited to identifying, customizing, and optimizing lifestyle (e.g., exercise, diet) and/or therapeutic (e.g., the particular drug or combination thereof) and/or medical intervention(s) (e.g., stent placement, angioplasty). The
10 same general principals also apply to the monitoring of CVD or CVD events, the severity of CVD or CVD events, and/or the response to lifestyle, therapeutic and/or medical intervention(s). Specifically, methods and compositions that include determining the methylation status of at least one CpG locus and/or at least one single nucleotide polymorphism (SNP) are described.

15 In one aspect, kits for determining methylation status of at least one CpG dinucleotide and/or 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 cg04988978, cg21161138,
20 cg12655112, cg03725309, cg12586707, and cg17901584 or at a second CpG dinucleotide in linkage disequilibrium with the first CpG dinucleotide at a GC locus selected from the group consisting of cg04988978, cg21161138, cg12655112, cg03725309, cg12586707, and cg17901584, 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/or
25 at least one second nucleic acid primer at least 8 nucleotides in length that is complementary to a DNA sequence of a first SNP selected from the group consisting of rs2869675, rs4376434, rs12129789, rs7585056, rs710987, rs4639796, rs1333048, rs12714414, rs942317, and rs1441433 or a second SNP in linkage disequilibrium with the first SNP selected from the group consisting of rs2869675, rs4376434, rs12129789, rs7585056,
30 rs710987, rs4639796, rs1333048, rs12714414, rs942317, and rs1441433, 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 embodiments, the at least one first nucleic acid primer detects the methylated CpG dinucleotide.

5 In some embodiments, 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 embodiments, 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.

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

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

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

In another aspect, methods of determining the presence of biomarkers associated with predicting, treating, managing and/or monitoring CVD in a biological sample from a patient is provided. Such methods typically include (a) providing a first portion of the biological
20 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 cg04988978, cg21161138, cg12655112, cg03725309, cg12586707, and
25 cg17901584, or a second CpG dinucleotide in linkage disequilibrium with the first CpG dinucleotide at a GC locus selected from the group consisting of cg04988978, cg21161138, cg12655112, cg03725309, cg12586707, and cg17901584, 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
30 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 rs2869675, rs4376434, rs12129789, rs7585056, rs710987, rs4639796, rs1333048, rs12714414, rs942317, and rs1441433 or a second SNP in linkage disequilibrium with the first SNP selected from the group consisting of rs2869675, rs4376434, rs12129789, rs7585056, rs710987, rs4639796, rs1333048, rs12714414, rs942317, and rs1441433, 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 cg04988978, cg21161138, cg12655112, cg03725309, cg12586707, and cg17901584, and the identity of the nucleotide at the first SNP selected from the group consisting of rs2869675, rs4376434, rs12129789, rs7585056, rs710987, rs4639796, rs1333048, rs12714414, rs942317, and rs1441433 or the second SNP in linkage disequilibrium with the first SNP are biomarkers associated with detecting CVD or estimating survival from CVD.

In some embodiments, the biological sample is blood or saliva.

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

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

In some embodiments, the window of incidence for detection, severity, managing and/or monitoring is three years, five years, or ten years.

In still a further aspect, methods of determining the presence of biomarkers in a biological sample from a subject, wherein the biomarkers are associated with detecting CVD, determining severity of CVD, estimating survival from CVD, identifying, customizing, and/or optimizing intervention(s) for CVD, managing CVD and/or monitoring CVD. 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 selected from rs2869675, rs4376434, rs12129789, rs7585056, rs710987, rs4639796, rs1333048, rs12714414, rs942317, and rs1441433 or from Appendix C and/or is a second SNP in linkage disequilibrium ($R > 0.3$) with a first SNP selected from rs2869675, rs4376434, rs12129789, rs7585056, rs710987, rs4639796, rs1333048, rs12714414, rs942317, and rs1441433 or 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 selected from cg04988978, cg21161138, cg12655112, cg03725309, cg12586707, and cg17901584 or from Appendix A and/or a CpG site collinear ($R > 0.3$) with a CpG site selected from cg04988978, cg21161138, cg12655112, cg03725309, cg12586707, and cg17901584 or from Appendix A to obtain methylation data; and (d) entering 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 embodiments, the at least one interaction effect is selected from the group consisting of a gene-environment interaction (SNPx CpG) effect, a gene-gene interaction (SNPx SNP) effect, and an environment-environment interaction (CpGxCpG) effect. In some embodiments, the at least one interaction effect is a gene-environment interaction effect (SNPx CpG) between a CpG site selected from cg04988978, cg21161138, cg12655112, cg03725309, cg12586707, and cg17901584 or from Appendix A or a CpG site that is collinear ($R > 0.3$) with a CpG site selected from cg04988978, cg21161138, cg12655112, cg03725309, cg12586707, and cg17901584 or from Appendix A and a SNP selected from rs2869675, rs4376434, rs12129789, rs7585056, rs710987, rs4639796, rs1333048, rs12714414, rs942317, and rs1441433 or from Appendix C or a SNP within moderate linkage disequilibrium ($R > 0.3$) from a SNP selected from rs2869675, rs4376434, rs12129789, rs7585056, rs710987, rs4639796, rs1333048, rs12714414, rs942317, and rs1441433 or from Appendix C. In some embodiments, the at least one interaction effect is an environment-environment interaction effect (CpGxCpG) between at least two CpG sites selected from cg04988978, cg21161138, cg12655112, cg03725309, cg12586707, and cg17901584 or from Appendix A.

In some embodiments, 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 selected from cg04988978, cg21161138, cg12655112, cg03725309, cg12586707, and cg17901584 or from Appendix A. In some embodiments, the at least one interaction effect is a gene-gene interaction effect (SNPx SNP)

between at least two SNPs selected from rs2869675, rs4376434, rs12129789, rs7585056, rs710987, rs4639796, rs1333048, rs12714414, rs942317, and rs1441433 or from Appendix C. In some embodiments, one or both of the at least two SNPs are collinear ($R > 0.3$) with one or both of the at least two SNPs selected from rs2869675, rs4376434, rs12129789, rs7585056, rs710987, rs4639796, rs1333048, rs12714414, rs942317, and rs1441433 or from Appendix C.

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

In another aspect, 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 selected from rs2869675, rs4376434, rs12129789, rs7585056, rs710987, rs4639796, rs1333048, rs12714414, rs942317, and rs1441433 or from Appendix C and/or is a second SNP in linkage disequilibrium ($R > 0.3$) with a first SNP selected from rs2869675, rs4376434, rs12129789, rs7585056, rs710987, rs4639796, rs1333048, rs12714414, rs942317, and rs1441433 or 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 selected from cg04988978, cg21161138, cg12655112, cg03725309, cg12586707, and cg17901584 or from Appendix A and/or a CpG site collinear ($R > 0.3$) with a CpG selected from cg04988978, cg21161138, cg12655112, cg03725309, cg12586707, and cg17901584 or 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 and/or methylation data.

In some embodiments, 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 and/or methylation data.

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

In some embodiments, dimensionality reductions (e.g., principal component analysis, partial least squares regression, etc.) can be used.

5 In yet another aspect, 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
10 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 selected from rs2869675, rs4376434, rs12129789, rs7585056, rs710987, rs4639796, rs1333048, rs12714414, rs942317, and rs1441433 or from Appendix C and/or is a second SNP in linkage disequilibrium ($R > 0.3$) with a first SNP selected from rs2869675, rs4376434, rs12129789, rs7585056, rs710987, rs4639796,
15 rs1333048, rs12714414, rs942317, and rs1441433 or 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 selected from cg04988978, cg21161138, cg12655112, cg03725309, cg12586707, and cg17901584 or from Appendix A and/or a CpG site collinear
20 ($R > 0.3$) with a CpG selected from cg04988978, cg21161138, cg12655112, cg03725309, cg12586707, and cg17901584 or from Appendix A to obtain methylation data.

In some embodiments, 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,
25 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, kits for determining methylation status of at least one CpG dinucleotide
30 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 cg04988978, cg21161138, cg12655112, cg03725309, cg12586707, and cg17901584 or at a second CpG dinucleotide in linkage disequilibrium with the first CpG dinucleotide at a GC locus selected from the group consisting of cg04988978, cg21161138, cg12655112, cg03725309, cg12586707, and cg17901584, 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.

In another aspect, methods of determining the presence of biomarkers in a biological sample from a subject, wherein the biomarkers are associated with detecting CVD, determining severity of CVD, estimating survival from CVD, identifying, customizing, and/or optimizing intervention(s) for CVD, managing CVD and/or monitoring CVD are provided. Such methods typically include (a) providing a biological sample from the subject at risk for or having CVD or CVD events, wherein nucleic acids from at least a portion of the biological sample are bisulfite converted; and (b) contacting the bisulfite converted nucleic acids 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 cg04988978, cg21161138, cg12655112, cg03725309, cg12586707, and cg17901584, or a second CpG dinucleotide in linkage disequilibrium with the first CpG dinucleotide at a GC locus selected from the group consisting of cg04988978, cg21161138, cg12655112, cg03725309, cg12586707, and cg17901584, wherein the linkage disequilibrium has a value of $R > 0.3$, wherein the first nucleic acid primer detects a methylated or unmethylated CpG dinucleotide, wherein the percentage of methylation of the CpG dinucleotide at the GC locus selected from the group consisting of cg04988978, cg21161138, cg12655112, cg03725309, cg12586707, and cg17901584 is associated with estimating survival of the subject.

In another aspect, methods of determining the presence of biomarkers in a biological sample from a subject, wherein the biomarkers are associated with detecting CVD, determining severity of CVD, estimating survival from CVD, identifying, customizing, and/or optimizing intervention(s) for CVD, managing CVD and/or monitoring CVD are provided. Such methods typically include (a) isolating nucleic acid sample from the subject sample; (b) bisulfite converting at least a portion of the nucleic acid and performing

methylation assessment on the bisulfite converted nucleic acid to determine the methylation status of at least one CpG site selected from cg04988978, cg21161138, cg12655112, cg03725309, cg12586707, and cg17901584 or from Appendix A and/or a CpG site collinear ($R > 0.3$) with a CpG site selected from cg04988978, cg21161138, cg12655112, cg03725309, cg12586707, and cg17901584 or from Appendix A to obtain methylation data; and (c) entering the methylation data from step (b) into an algorithm that accounts for at least one CpG main effect, wherein the algorithm is a machine learning algorithm capable of accounting for linear and non-linear effects.

In still another aspect, systems for determining methylation status of at least one CpG dinucleotide are provided. Such systems typically include a nucleic acid isolation module configured to isolate a nucleic acid sample from a subject sample; a methylation assay module configured to bisulfite convert the nucleic acid in at least a portion of the nucleic acid and perform a methylation assessment on the bisulfite converted nucleic acid to determine methylation status of at least one CpG site selected from cg04988978, cg21161138, cg12655112, cg03725309, cg12586707, and cg17901584 or from Appendix A and/or a CpG site collinear ($R > 0.3$) with a CpG site selected from cg04988978, cg21161138, cg12655112, cg03725309, cg12586707, and cg17901584 or from Appendix A to obtain methylation data; and an identification system configured to account for at least one CpG main effect based on the methylation data.

In yet another aspect, a non-transitory computer-readable medium storing instructions executable by a processing device to perform operations is provided. Such a computer-readable medium typically includes accounting for at least one CpG main effect based on methylation data, wherein: the methylation data is based on a methylation assay on a bisulfite converted nucleic acid in at least a portion of a nucleic acid sample to detect methylation status of at least one CpG site selected from cg04988978, cg21161138, cg12655112, cg03725309, cg12586707, and cg17901584 or from Appendix A and/or a CpG site collinear ($R > 0.3$) with a CpG site selected from cg04988978, cg21161138, cg12655112, cg03725309, cg12586707, and cg17901584 or from Appendix A to obtain methylation data.

The integrated genetic-epigenetic model described herein provides several advantages and benefits. For example,

- Earlier Detection: PrecisionCHD can detect molecular changes that may or may not precede the development of clinical symptoms or disease. This means patients can be identified with coronary heart disease before or after they develop symptoms, allowing for earlier interventions and better outcomes.
- 5 • Actionable Clinical Intelligence™: The test may or may not be coupled to a provider-facing Actionable Clinical Intelligence platform (see, for example, US Application No. 63/488,463, incorporated herein by reference) that maps each patient's molecular markers to key drivers of coronary heart disease, allowing for tailored recommendations for lifestyle modifications, medical interventions, estimating effectiveness of lifestyle modifications or medical interventions, monitoring and secondary testing.
- 10 • Personalized Intervention Selection: The test can be used to select one or more interventions such as lifestyle modification, therapeutic interventions and medical interventions for each patient or a group of patients at one or more time points.
- 15 • Personalized Intervention Optimization: The test can be used to optimize one or more interventions such as lifestyle modification, therapeutic interventions and medical interventions for each patient or a group of patients at one or more time points.
- 20 • Personalized Intervention Evaluation: The test can be used to assess the effectiveness of interventions such as lifestyle modification, therapeutic interventions, and medical interventions for each patient or a group of patients by continually monitoring CVD, CVD events, or severity of CVD.
- 25 • More Comprehensive Assessment: The test evaluates and integrates robust genetic and epigenetic biomarkers simultaneously, providing a more comprehensive assessment of CVD status.
- 30 • Discover New Pathways: The approach can be used to discover new, previously unknown biological pathways for risk assessment, detection, intervention (e.g., lifestyle, therapeutic, medical), management and monitoring of CVD. The pathway(s) and biomarker(s) also can be used for the discovery, development and validation of novel biopharmaceuticals for the assessment and management of cardiovascular disease. The approach described herein can be used to identify new biomarker(s) (e.g., methylation, SNP, protein, etc.) for new drug development or the ability for

targeted treatment such as gene editing. The approach described herein also can be used to discover biomarker(s) to select the most effective drug for a particular individual (e.g., statin vs. beta blocker), the most effective drug type (e.g., hydrophilic statin vs. lipophilic statin), changes in lifestyle, medical interventions, or combinations thereof. In addition, the approach described herein can be used to optimize the use of a therapeutic (e.g., dosing, regimen, drug combinations) or to identify lifestyle changes that would have the most effect.

- Non-Invasive: The test only requires a simple biomaterial collection (e.g., blood or saliva sample), making it a non-invasive and convenient alternative to more invasive diagnostic tests such as angiograms.
- Accessible: The test can be administered remotely via a lancet-based sample collection kit that can be sent to the patient's home upon test order, thereby increasing and democratizing access to CVD diagnostic tests. Or biomaterial can be collected in-provider settings via a vacutainer-based sample collection. The biomaterial in the form of a saliva sample can be collected remotely or in-provider settings.
- Cost-effective and Timely: PrecisionCHD provides clinicians with a timely and cost-effective coronary heart disease test. PrecisionCHD is a fraction of the cost of other heart disease tests.
- Survival Estimates: PrecisionCHD or PrecisionCHD-Epi can provide survival estimates for those individuals that have already been determined to have CVD or are considered at risk of developing CVD.

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 of matter, suitable methods and materials are described below. In addition, the materials, methods, and examples are illustrative only and not intended to be limited to predicting incident CHD. All publications, patent applications, patents, and other references mentioned herein are incorporated by reference in their entirety.

DESCRIPTION OF DRAWINGS

FIG. 1 is a block diagram of an example cardiovascular disease classification and monitoring system.

FIG. 2 is a flow diagram of an example process for cardiovascular disease
5 classification and monitoring.

FIG. 3 is a block diagram of example computing devices.

FIG. 4 is a schematic showing potential therapeutic approaches for patients that have a positive signal for CHD and/or associated co-morbidities using the methods described herein.

10 FIGs. 5A - 5C are plots showing survival (yes or no) during follow-up to methylation for markers cg04988978 (FIG. 5A), cg21161138 (FIG. 5B) and cg12655112 (FIG. 5C).

FIGs. 6A - 6C are plots showing the relationship between days to death and methylation at markers cg04988978 (FIG. 6A), cg21161138 (FIG. 6B) and cg12655112 (FIG. 6C).

15 DETAILED DESCRIPTION

Recent prediction strategies have taken advantage of the rapid advancements in assessing genome-wide genetic or transcriptional variation. Though each of these approaches have had some success, their clinical impact has been limited. In particular, those relying only on genetic information have a clear ceiling in predictive capacity, are potentially
20 sensitive to ethnic stratification, and, because genotype is static, cannot be used to monitor changes in disease status.

Recent advances in genome-wide epigenetic profiling techniques have raised the possibility that DNA methylation assessments of peripheral blood DNA may serve as a mechanism for more accurate prediction of cardiovascular disease or mortality. Prediction
25 models that only account for epigenetic signatures, however, fail to account for confounding genetic variation, which affects the vast majority of the environmentally responsive methylome. This may result in models that lack robustness with respect to generalizability, especially in different ethnic groups.

As a result, we have developed a highly sensitive, clinically implementable integrated
30 genetic-epigenetic tool capable of identifying those at risk of or with cardiovascular disease

(e.g., having a heart attack or sudden cardiac death). As shown herein, the methylation status of one or more particular CpG dinucleotides in combination with the genotype at one or more particular loci (e.g., CH3xSNP) can be used to predict cardiovascular disease (CVD) including coronary heart disease (CHD). We also have developed a highly sensitive tool capable of estimating the survivability of those who are at risk of developing CVD, those who are at risk of having a CVD event, or those who have already been identified as having CVD.

As described herein, biomarkers can be used in the diagnosis and prognosis of cardiovascular diseases and events. The terms “marker” and “biomarker” can be used interchangeably. As used herein, a biomarker generally refers to a measurable or detectable biological moiety (e.g., the presence or amount of a protein, a genetic (e.g., polymorphism), epigenetic (e.g., methylation), and/or histological component). As described in more detail below, the biomarkers used herein typically are associated with cardiovascular disease.

As used herein, “patient,” “subject” and “individual” may be used interchangeably.

DNA Methylation

DNA does not exist as naked molecules in the cell. For example, DNA is associated with proteins called histones to form a complex substance known as chromatin. Chemical modifications of the DNA or the histones alter the structure of the chromatin without changing the nucleotide sequence of the DNA. Such modifications are described as “epigenetic” modifications of the DNA. Changes to the structure of the chromatin can have a profound influence on gene expression. If the chromatin is condensed, factors involved in gene expression may not have access to the DNA, and the genes will be switched off. Conversely, if the chromatin is “open,” the genes can be switched on. Some important forms of epigenetic modification are DNA methylation and histone deacetylation.

DNA methylation is a chemical modification of the DNA molecule itself and is carried out by an enzyme called DNA methyltransferase. Methylation can directly switch off gene expression by preventing transcription factors binding to promoters. A more general effect is the attraction of methyl-binding domain (MBD) proteins. These are associated with further enzymes called histone deacetylases (HDACs), which function to chemically modify histones and change chromatin structure. Chromatin-containing acetylated histones are open

and accessible to transcription factors, and the genes are potentially active. Histone deacetylation causes the condensation of chromatin, making it inaccessible to transcription factors and causing the silencing of genes.

CpG islands are short stretches of DNA in which the frequency of the CpG sequence is higher than other regions. The “p” in the term CpG indicates that cytosine (“C”) and guanine (“G”) are connected by a phosphodiester bond. CpG islands are often located around promoters of housekeeping genes and many regulated genes. At these locations, the CG sequence in active genes are oftentimes not methylated. By contrast, the CG sequences in inactive genes are usually methylated to suppress their expression.

As used herein, the term “methylation status” means the determination whether a certain target DNA, such as a CpG dinucleotide, is methylated or is unmethylated. As used herein, the term “CpG dinucleotide repeat motif” means a series of two or more CpG dinucleotides positioned in a DNA sequence.

About 56% of human genes and 47% of mouse genes are associated with CpG islands. Often, CpG islands overlap the promoter and extend about 1000 base pairs downstream into the transcription unit. Identification of potential CpG islands during sequence analysis helps to define the extreme 5' ends of genes, something that is notoriously difficult with cDNA-based approaches. The methylation of a CpG island can be determined by a skilled artisan using any method suitable to determine such methylation. For example, the skilled artisan can use a bisulfite reaction-based method for determining such methylation.

The present disclosure provides methods to determine the nucleic acid methylation of one or more loci in a subject in order to identify subjects having CVD.

Linkage refers to the phenomenon that DNA sequences which are close together in the genome have a tendency to be inherited together. Two sequences may be linked because of some selective advantage of co-inheritance. More typically, however, two sequences are co-inherited because of the relative infrequency with which meiotic recombination events occur within the region between the two sequences. The co-inherited sequences are said to be in “linkage disequilibrium” with one another because, in a given population, they tend to either both occur together or else not occur at all in any particular member of the population. Indeed, where multiple sequences in a given chromosomal region are found to be in linkage

disequilibrium with one another, they define a quasi-stable "haplotype." In contrast, recombination events occurring between two loci cause them to become separated onto distinct homologous chromosomes. If meiotic recombination between two physically linked sequences occurs frequently enough, the two sequences will appear to segregate independently and are said to be in linkage equilibrium.

It would be understood that linkage disequilibrium can be quantitated (using, for example, the Pearson correlation (R) or co-inheritance of alleles (D')). For example, a low level of linkage can be reflected in a correlation (e.g., R value) of about 0.1 or less, a moderate level of linkage is reflected in a R value of about 0.3, while a high level of linkage is reflected in a R value of 0.5 or greater. It also would be understood that, when referring to methylation (i.e., CpG sites), collinearity (with an R value) is used as a determination of the linear strength of the association between two CpGs (e.g., a low level of collinearity can be reflected by an R value of about 0.1 or less; a moderate level of collinearity can be reflected by an R value of about 0.3; and a high level of collinearity can be reflected by an R value of about 0.5 or greater).

In particular, in certain embodiments of the disclosure, the methods may be practiced as follows. A sample, such as a blood sample, is taken from a subject. In certain embodiments, a single cell type, e.g., lymphocytes, basophils, or monocytes isolated from the blood, may be isolated for further testing. The DNA is harvested from the sample and examined to determine the methylation of one or more loci. For example, the DNA of interest can be treated with bisulfite to deaminate unmethylated cytosine residues to uracil. Since uracil base pairs with adenosine, thymidines are incorporated into subsequent DNA strands in the place of unmethylated cytosine residues during subsequent PCR amplifications. Next, the target sequence is amplified by PCR, and probed with a loci-specific probe. Depending on the particular sequence of the probe used, only the methylated or unmethylated DNA will bind to the probe.

Methods of determining the subject nucleic acid profile are well known to a skilled artisan and include any of the well-known detection methods. Various PCR methods are described, for example, in *PCR Primer: A Laboratory Manual*, Dieffenbach & Dveksler, Eds., Cold Spring Harbor Laboratory Press, 1995. Other methods include, but are not limited to, nucleic acid quantification, restriction enzyme digestion, DNA sequencing, hybridization

technologies, such as Southern Blotting, amplification methods such as Ligase Chain Reaction (LCR), Nucleic Acid Sequence Based Amplification (NASBA), Self-sustained Sequence Replication (SSR or 3SR), Strand Displacement Amplification (SDA), and Transcription Mediated Amplification (TMA), Quantitative PCR (qPCR), digital PCR (dPCR) (e.g., digital droplet PCR (ddPCR)) or other DNA analyses, as well as RT-PCR, in vitro translation, Northern blotting, and other RNA analyses. In another embodiment, hybridization on a microarray is used.

Single Nucleotide Polymorphism (SNP)

Traditional methods for the screening of heritable diseases have depended on either the identification of abnormal gene products (e.g., sickle cell anemia) or an abnormal phenotype (e.g., mental retardation). With the development of simple and inexpensive genetic screening methodology, it is now possible to identify polymorphisms that indicate a propensity to develop disease, even when the disease is of polygenic origin.

Single nucleotide polymorphism (SNP) genotyping measures genetic variations of SNPs between members of a species. A SNP is a single base pair change at a specific locus, usually consisting of two alleles (where the rare allele frequency is >1%). SNPs are very common. Because SNPs are conserved during evolution, they have been proposed as markers for use in quantitative trait loci (QTL) analysis and in association studies in place of microsatellites. Many different SNP genotyping methods are known, including hybridization-based methods (such as Dynamic allele-specific hybridization, molecular beacons, and SNP microarrays) enzyme-based methods (including restriction fragment length polymorphism, PCR-based methods, flap endonuclease, primer extension, 5'-nuclease, and oligonucleotide ligation assay), other post-amplification methods based on physical properties of DNA (such as single strand conformation polymorphism, temperature gradient gel electrophoresis, denaturing high performance liquid chromatography, high-resolution melting of the entire amplicon, use of DNA mismatch-binding proteins, SNPLEX and surveyor nuclease assay), and sequencing (such as "next generation" sequencing). See, e.g., US Patent No. 7,972,779.

A plurality of alleles at a locus can arise from one or more polymorphisms in a region of a gene that encodes a polypeptide or in a regulatory control sequence that affects expression of the polypeptide, such as a promoter or polyadenylation sequence.

Alternatively, alleles can arise from one or more polymorphisms at a locus distal to a gene that encodes a polypeptide or in a regulatory control sequence. A polymorphism can affect a polypeptide at a transcriptional or a translational level (e.g., a polypeptide's transcription rate, translation rate, degradation rate, and/or activity). Allelic differences can be characterized in a sample from a single subject or from a plurality of subjects using methods that are known to a skilled artisan. Such methods can include, but are not limited to, measuring the potential for a polynucleotide sequence to be expressed and/or measuring an amount of an encoded polypeptide. Methods are available that can detect proteins or nucleic acids directly or indirectly, and assay methods are specifically contemplated to include screening for the presence of particular sequences or structures of nucleic acids or polypeptides using, e.g., any of various known microarray technologies.

It will be fully appreciated by the skilled artisan that the allele need not have previously been shown to have had any link or association with the disorder phenotype. Instead, an allele and a pathogenic environmental risk factor can interact to predict a predisposition to a disorder phenotype even when neither the allele nor the risk factor bears any direct relation to the disorder phenotype.

Genetic screening (also called genotyping or molecular screening) can be broadly defined as testing to determine if a subject has mutations (or alleles or polymorphisms) that either cause a disease state or are "linked" to the mutation causing a disease state. Linkage refers to the phenomenon that DNA sequences which are close together in the genome have a tendency to be inherited together. Two sequences may be linked because of some selective advantage of co-inheritance. More typically, however, two polymorphic sequences are co-inherited because of the relative infrequency with which meiotic recombination events occur within the region between the two polymorphisms. The co-inherited polymorphic alleles are said to be in "linkage disequilibrium" with one another because, in a given population, they tend to either both occur together or else not occur at all in any particular member of the population. Indeed, where multiple polymorphisms in a given chromosomal region are found to be in linkage disequilibrium with one another, they define a quasi-stable genetic "haplotype." In contrast, recombination events occurring between two polymorphic loci cause them to become separated onto distinct homologous chromosomes. If meiotic recombination between two physically linked polymorphisms occurs frequently enough, the

two polymorphisms will appear to segregate independently and are said to be in linkage equilibrium.

It would be understood that linkage disequilibrium can be quantitated (using, for example, the Pearson correlation (R) or co-inheritance of alleles (D')). For example, a low
5 level of linkage can be reflected in a correlation (e.g., R value) of about 0.1 or less, a moderate level of linkage is reflected in a R value of about 0.3, while a high level of linkage is reflected in a R value of 0.5 or greater.

While the frequency of meiotic recombination between two markers is generally proportional to the physical distance between them on the chromosome, the occurrence of
10 "hot spots" as well as regions of repressed chromosomal recombination can result in discrepancies between the physical and recombinatorial distance between two markers. Thus, in certain chromosomal regions, multiple polymorphic loci spanning a broad chromosomal domain may be in linkage disequilibrium with one another, and thereby define a broad-spanning genetic haplotype. Furthermore, where a disease-causing mutation is found within
15 or in linkage with this haplotype, one or more polymorphic alleles of the haplotype can be used as a diagnostic or prognostic indicator of the likelihood of developing the disease. This association between otherwise benign polymorphisms and a disease-causing polymorphism occurs if the disease mutation arose in the recent past, so that sufficient time has not elapsed for equilibrium to be achieved through recombination events. Therefore, identification of a
20 haplotype that spans or is linked to a disease-causing mutational change serves as a predictive measure of an individual's likelihood of having inherited that disease-causing mutation. Such prognostic or diagnostic procedures can be utilized without necessitating the identification and isolation of the actual disease-causing lesion. This is significant because the precise determination of the molecular defect involved in a disease process can be
25 difficult and laborious, especially in the case of multifactorial diseases.

The statistical correlation between a disorder and a polymorphism does not necessarily indicate that the polymorphism directly causes the disorder. Rather the correlated polymorphism may be a benign allelic variant which is linked to (i.e., in linkage disequilibrium with) a disorder-causing mutation that has occurred in the recent evolutionary
30 past, so that sufficient time has not elapsed for equilibrium to be achieved through recombination events in the intervening chromosomal segment. Thus, for the purposes of

diagnostic and prognostic assays for a particular disease, detection of a polymorphic allele associated with that disease can be utilized without consideration of whether the polymorphism is directly involved in the etiology of the disease. Furthermore, where a given benign polymorphic locus is in linkage disequilibrium with an apparent disease-causing polymorphic locus, still other polymorphic loci which are in linkage disequilibrium with the benign polymorphic locus are also likely to be in linkage disequilibrium with the disease-causing polymorphic locus. Thus, these other polymorphic loci will also be prognostic or diagnostic of the likelihood of having inherited the disease-causing polymorphic locus. A broad-spanning haplotype (describing the typical pattern of co-inheritance of alleles of a set of linked polymorphic markers) can be targeted for diagnostic purposes once an association has been drawn between a particular disease or condition and a corresponding haplotype. Thus, the determination of an individual's likelihood for developing a particular disease or condition can be made by characterizing one or more disease-associated polymorphic alleles (or even one or more disease-associated haplotypes) without necessarily determining or characterizing the causative genetic variation.

Many methods are available for detecting specific alleles at polymorphic loci. Certain methods for detecting a specific polymorphic allele will depend, in part, upon the molecular nature of the polymorphism. For example, the various allelic forms of the polymorphic locus may differ by a single base-pair of the DNA. Such single nucleotide polymorphisms (or SNPs) are major contributors to genetic variation, comprising some 80% of all known polymorphisms, and their density in the genome is estimated to be on average 1 per 1,000 base pairs. SNPs are most frequently bi-allelic, or occurring in only two different forms (although up to four different forms of an SNP, corresponding to the four different nucleotide bases occurring in DNA, are theoretically possible). Nevertheless, SNPs are mutationally more stable than other polymorphisms, making them suitable for association studies in which linkage disequilibrium between markers and an unknown variant is used to map disease-causing mutations. In addition, because SNPs typically have only two alleles, they can be genotyped by a simple plus / minus assay rather than a length measurement, making them more amenable to automation.

In one embodiment, allelic profiling can be accomplished using a nucleic acid microarray. The genetic testing field is rapidly evolving and, as such, the skilled artisan will

appreciate that a wide range of profiling tests exist, and will be developed, to determine the allelic profile of individuals in accord with the disclosure.

Nucleic Acids and Polypeptides

5 As described herein, the methods provided in this disclosure rely upon features contained within the nucleic acid of an individual, subject or patient. The term “nucleic acid” refers to deoxyribonucleotides or ribonucleotides and polymers thereof in either single- or double-stranded form, made of monomers (nucleotides) containing a sugar, phosphate and a base that is either a purine or pyrimidine. Unless specifically limited, the term encompasses
10 nucleic acids containing known analogs of natural nucleotides that have similar binding properties as the reference nucleic acid and are metabolized in a manner similar to naturally occurring nucleotides. Unless otherwise indicated, a particular nucleic acid sequence also encompasses conservatively modified variants thereof (*e.g.*, degenerate codon substitutions) and complementary sequences, as well as the sequence explicitly indicated. Specifically,
15 degenerate codon substitutions may be achieved by generating sequences in which the third position of one or more selected (or all) codons is substituted with mixed-base and/or deoxyinosine residues. The terms “nucleic acid,” “nucleic acid molecule,” or “polynucleotide” are used interchangeably and may also be used interchangeably with gene, cDNA, DNA and/or RNA encoded by a gene.

20 The term “nucleotide sequence” refers to a polymer of DNA or RNA which can be single-stranded or double-stranded, optionally containing synthetic, non-natural or altered nucleotide bases capable of incorporation into DNA or RNA polymers. A DNA molecule or polynucleotide is a polymer of deoxyribonucleotides (A, G, C, and T), and an RNA molecule or polynucleotide is a polymer of ribonucleotides (A, G, C and U).

25 A “gene,” for the purposes of the present disclosure, includes a DNA region encoding a gene product, as well as all DNA regions, which regulate the production of the gene product, whether or not such regulatory sequences are adjacent to coding and/or transcribed sequences. The term “gene” is used broadly to refer to any segment of nucleic acid associated with a biological function. Genes include coding sequences and/or the regulatory sequences
30 required for their expression. Accordingly, a gene includes, but is not necessarily limited to, promoter sequences, terminators, translational regulatory sequences such as ribosome

binding sites and internal ribosome entry sites, enhancers, silencers, insulators, boundary elements, replication origins, matrix attachment sites and locus control regions. For example, “gene” refers to a nucleic acid fragment that expresses mRNA, functional RNA, or specific protein, including regulatory sequences. “Functional RNA” refers to sense RNA, antisense RNA, ribozyme RNA, siRNA, or other RNA that may not be translated but yet has an effect on at least one cellular process. “Genes” also include non-expressed DNA segments that, for example, form recognition sequences for other proteins. “Genes” can be obtained from a variety of sources, including cloning from a source of interest or synthesizing from known or predicted sequence information, and may include sequences designed to have desired parameters.

“Gene expression” refers to the conversion of the information, contained in a gene, into a gene product. It refers to the transcription and/or translation of an endogenous gene, heterologous gene or nucleic acid segment, or a transgene in cells. In addition, expression refers to the transcription and stable accumulation of sense (mRNA) or functional RNA. Expression may also refer to the production of protein. The term “altered level of expression” refers to the level of expression in transgenic cells or organisms that differs from that of normal or untransformed cells or organisms.

A gene product can be the transcriptional product of a gene (*e.g.*, mRNA, tRNA, rRNA, antisense RNA, ribozyme, structural RNA or any other type of RNA) or a protein produced by translation of an mRNA. Gene products also include RNAs that are modified, by processes such as capping, polyadenylation, methylation, and editing, and proteins modified by, for example, methylation, acetylation, phosphorylation, ubiquitination, ADP-ribosylation, myristilation, and glycosylation. The term “RNA transcript” refers to the product resulting from RNA polymerase-catalyzed transcription of a DNA sequence. When the RNA transcript is a complementary copy of the DNA sequence, it is referred to as the primary transcript; a RNA sequence derived from post-transcriptional processing of the primary transcript is referred to as the mature RNA. “Messenger RNA” (mRNA) refers to the RNA that lacks introns and that can be translated into protein by the cell. “cDNA” refers to a single- or a double-stranded DNA that is complementary to and derived from mRNA. “Functional RNA” refers to sense RNA, antisense RNA, ribozyme RNA, siRNA, or other RNA that may not be translated but yet has an effect on at least one cellular process.

A "coding sequence" or a sequence that "encodes" a polypeptide is a nucleic acid molecule that is transcribed (in the case of DNA) and/or translated (in the case of mRNA) into a polypeptide *in vivo* when placed under the control of appropriate regulatory sequences. The boundaries of the coding sequence are determined by a start codon at the 5' (amino) terminus and a translation stop codon at the 3' (carboxy) terminus. A coding sequence can include, but is not limited to, cDNA from viral, prokaryotic or eukaryotic mRNA, genomic DNA sequences from viral (e.g., DNA viruses and retroviruses) or prokaryotic DNA, and synthetic DNA sequences. A transcription termination sequence can be located 3' to the coding sequence.

"Regulatory sequences" and "suitable regulatory sequences" each refer to nucleotide sequences located upstream (5' non-coding sequences), within, or downstream (3' non-coding sequences) of a coding sequence, and which influence the transcription, RNA processing or stability, or translation of the associated coding sequence. Regulatory sequences include enhancers, promoters, translation leader sequences, introns, and polyadenylation signal sequences. They include natural and synthetic sequences as well as sequences that may be a combination of synthetic and natural sequences.

Certain embodiments of the disclosure encompass isolated or substantially purified nucleic acid compositions. In the context of the present disclosure, an "isolated" or "purified" DNA molecule or RNA molecule is a DNA molecule or RNA molecule that exists apart from its native environment and is, therefore, not a product of nature. An isolated DNA molecule or RNA molecule may exist in a purified form or may exist in a non-native environment such as, for example, a transgenic host cell. For example, an "isolated" or "purified" nucleic acid molecule is substantially free of other cellular material, or culture medium when produced by recombinant techniques, or substantially free of chemical precursors or other chemicals when chemically synthesized. In one embodiment, an "isolated" nucleic acid is free of sequences that naturally flank the nucleic acid (i.e., sequences located at the 5' and 3' ends of the nucleic acid) in the genomic DNA of the organism from which the nucleic acid is derived.

By "fragment" is intended a polypeptide consisting of only a part of the intact full-length polypeptide sequence and structure. The fragment can include a C-terminal deletion, an N-terminal deletion, and/or an internal deletion of the native polypeptide. A fragment of a protein will generally include at least about 5-100 contiguous amino acid residues of the full-

length molecule (e.g., at least about 15-25 contiguous amino acid residues of the full-length molecule, at least about 20-50 or more contiguous amino acid residues of the full-length molecule, or any integer between 5 amino acids and the full-length sequence).

5 “Naturally occurring” is used to describe a composition that can be found in nature as distinct from being artificially produced. For example, a nucleotide sequence present in an organism, which can be isolated from a source in nature and which has not been intentionally modified by a person in the laboratory, is naturally occurring.

A “5’ non-coding sequence” refers to a nucleotide sequence located 5’ (upstream) to the coding sequence. 5’ non-coding sequences are present in the fully processed mRNA
10 upstream of the initiation codon and may affect processing of the primary transcript to mRNA, mRNA stability or translation efficiency. A “3’ non-coding sequence” refers to nucleotide sequences located 3’ (downstream) to a coding sequence and may include polyadenylation signal sequences and other sequences encoding regulatory signals capable of affecting mRNA processing or gene expression.

15 A “promoter” refers to a nucleotide sequence, usually upstream (5’) to its coding sequence, which directs and/or controls the expression of the coding sequence by providing the recognition for RNA polymerase and other factors required for proper transcription. “Promoter” can include a minimal promoter that is a short DNA sequence comprised of a TATA-box and other sequences that serve to specify the site of transcription initiation, to
20 which regulatory elements are added for control of expression. “Promoter” also can refer to a nucleotide sequence that includes a minimal promoter plus one or more regulatory elements (e.g., enhancers) that are capable of controlling the expression of a coding sequence or functional RNA. Promoters may be derived in their entirety from a native sequence or be composed of different elements derived from different promoters found in nature, or even be
25 comprised of synthetic DNA sequences. A promoter may also contain DNA sequences that are involved in the binding of protein factors that control the effectiveness of transcription initiation in response to physiological or developmental conditions. “Constitutive expression” refers to expression using a constitutive promoter. “Conditional” and “regulated expression” refer to expression controlled by a regulated promoter.

30 An “enhancer” is a DNA sequence that can stimulate promoter activity. An enhancer may be an innate element of the promoter or a heterologous element inserted to enhance the

level or tissue specificity of a promoter. Enhancers often are capable of operating in both orientations and are capable of functioning even when moved either upstream or downstream from the promoter. Both enhancers and other regulatory elements within a promoter bind sequence-specific DNA-binding proteins that mediate their effects.

5 “Operably linked” refers to the association of nucleic acid sequences on a single nucleic acid fragment so that the function of one of the sequences is affected by another. For example, a regulatory DNA sequence is said to be “operably linked to” or “associated with” a DNA sequence that codes for an RNA or a polypeptide if the two sequences are situated such that the regulatory DNA sequence affects expression of the coding DNA sequence (i.e., that
10 the coding sequence or functional RNA is under the transcriptional control of the promoter). Coding sequences can be operably linked to regulatory sequences in sense or antisense orientation.

 “Expression” refers to the transcription and/or translation of an endogenous gene, heterologous gene or nucleic acid segment, or a transgene in cells. In addition, expression
15 refers to the transcription and stable accumulation of sense (mRNA) or functional RNA. Expression may also refer to the production of protein. The term “altered level of expression” refers to a level of expression in cells or organisms that differs from that of normal cells or organisms.

 For sequence comparison, typically one sequence acts as a reference sequence to
20 which test sequences are compared. When using a sequence comparison algorithm, test and reference sequences are input into a computer, and sequence algorithm program parameters are designated. The sequence comparison algorithm then calculates the percent sequence identity for the test sequence(s) relative to the reference sequence, based on the designated algorithm parameters.

25 The following terms are used to describe the sequence relationships between two or more nucleic acids or polynucleotides: (a) “reference sequence,” (b) “comparison window,” (c) “sequence identity,” (d) “percentage of sequence identity,” and (e) “as is for sequence comparison. A reference sequence may be a subset or the substantial identity.” As used herein, “reference sequence” is a defined sequence used as a entirety of a specified
30 sequence; for example, as a segment of a full-length cDNA or gene sequence, or the complete cDNA or gene sequence. As used herein, “comparison window” makes reference to

a contiguous and specified segment of a polynucleotide sequence, wherein the polynucleotide sequence in the comparison window may comprise additions or deletions (i.e., gaps) compared to the reference sequence (which does not comprise additions or deletions) for optimal alignment of the two sequences. Generally, the comparison window is at least 20 contiguous nucleotides in length, and optionally can be 30, 40, 50, 100, or longer. Those of skill in the art understand that, to avoid a high similarity to a reference sequence due to inclusion of gaps in the polynucleotide sequence, a gap penalty is typically introduced and is subtracted from the number of matches.

Methods of alignment of sequences for comparison are well-known in the art. Thus, the determination of percent identity between any two sequences can be accomplished using a mathematical algorithm. Non-limiting examples of such mathematical algorithms are the algorithm of Myers and Miller (Myers and Miller, CABIOS, 4, 11 (1988)); the local homology algorithm of Smith et al. (Smith et al., Adv. Appl. Math., 2, 482 (1981)); the homology alignment algorithm of Needleman and Wunsch (Needleman and Wunsch, JMB, 48, 443 (1970)); the search-for-similarity-method of Pearson and Lipman (Pearson and Lipman, Proc. Natl. Acad. Sci. USA, 85, 2444 (1988)); the algorithm of Karlin and Altschul (Karlin and Altschul, Proc. Natl. Acad. Sci. USA, 87, 2264 (1990)), modified as in Karlin and Altschul (Karlin and Altschul, Proc. Natl. Acad. Sci. USA 90, 5873 (1993)).

Computer implementations of these mathematical algorithms can be utilized for comparison of sequences to determine sequence identity. Such implementations include but are not limited to: CLUSTAL in the PC/Gene program (available from Intelligenetics, Mountain View, Calif.); the ALIGN program (Version 2.0) and GAP, BESTFIT, BLAST, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Version 8 (available from Genetics Computer Group (GCG), 575 Science Drive, Madison, Wis., USA).

Alignments using these programs can be performed using the default parameters. The CLUSTAL program is well described by Higgins et al. (Higgins et al., CABIOS, 5, 151 (1989)); Corpet et al. (Corpet et al., Nucl. Acids Res., 16, 10881 (1988)); Huang et al. (Huang et al., CABIOS, 8, 155 (1992)); and Pearson et al. (Pearson et al., Meth. Mol. Biol., 24, 307 (1994)). The ALIGN program is based on the algorithm of Myers and Miller, *supra*. The BLAST programs of Altschul et al. (Altschul et al., J. Mol. Biol., 215, 403 (1990)) are based on the algorithm of Karlin and Altschul, *supra*.

Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information. This algorithm involves first identifying high scoring sequence pairs (HSPs) by identifying short words of length "W" in the query sequence, which either match or satisfy some positive-valued threshold score T when aligned with a word of the same length in a database sequence. "T" is referred to as the neighborhood word score threshold. These initial neighborhood word hits act as seeds for initiating searches to find longer HSPs containing them. The word hits are then extended in both directions along each sequence for as far as the cumulative alignment score can be increased. Cumulative scores are calculated using, for nucleotide sequences, the parameters "M" (reward score for a pair of matching residues; always >0) and "N" (penalty score for mismatching residues; always <0), and for amino acid sequences, a scoring matrix is used to calculate the cumulative score. Extension of the word hits in each direction are halted when the cumulative alignment score falls off by the quantity "X" from its maximum achieved value, the cumulative score goes to zero or below due to the accumulation of one or more negative-scoring residue alignments, or the end of either sequence is reached.

In addition to calculating percent sequence identity, the BLAST algorithm also performs a statistical analysis of the similarity between two sequences. One measure of similarity provided by the BLAST algorithm is the smallest sum probability (P(N)), which provides an indication of the probability by which a match between two nucleotide or amino acid sequences would occur by chance. For example, a test nucleic acid sequence is considered similar to a reference sequence if the smallest sum probability in a comparison of the test nucleic acid sequence to the reference nucleic acid sequence is less than about 0.1, less than about 0.01, or even less than about 0.001.

To obtain gapped alignments for comparison purposes, Gapped BLAST (in BLAST 2.0) can be utilized. Alternatively, PSI-BLAST (in BLAST 2.0) can be used to perform an iterated search that detects distant relationships between molecules. When utilizing BLAST, Gapped BLAST, PSI-BLAST, the default parameters of the respective programs (e.g., BLASTN for nucleotide sequences, BLASTX for proteins) can be used. The BLASTN program (for nucleotide sequences) uses as defaults a word length (W) of 11, an expectation (E) of 10, a cutoff of 100, M=5, N=-4, and a comparison of both strands. For amino acid sequences, the BLASTP program uses as defaults a word length (W) of 3, an expectation (E)

of 10, and the BLOSUM62 scoring matrix. Alignment may also be performed manually by inspection.

For purposes of the present disclosure, comparison of nucleotide sequences for determination of percent sequence identity to the promoter sequences disclosed herein may be made using the BlastN program (version 1.4.7 or later) with its default parameters or any equivalent program. By “equivalent program” is intended any sequence comparison program that, for any two sequences in question, generates an alignment having identical nucleotide or amino acid residue matches and an identical percent sequence identity when compared to the corresponding alignment generated by the program.

As used herein, “sequence identity” or “identity” in the context of two nucleic acid or polypeptide sequences makes reference to a specified percentage of residues in the two sequences that are the same when aligned for maximum correspondence over a specified comparison window, as measured by sequence comparison algorithms or by visual inspection. When percentage of sequence identity is used in reference to proteins it is recognized that residue positions which are not identical often differ by conservative amino acid substitutions, where amino acid residues are substituted for other amino acid residues with similar chemical properties (e.g., charge or hydrophobicity) and, therefore, do not change the functional properties of the molecule. When sequences differ in conservative substitutions, the percent sequence identity may be adjusted upwards to correct for the conservative nature of the substitution. Sequences that differ by such conservative substitutions are said to have “sequence similarity” or “similarity.” Means for making this adjustment are well known to those of skill in the art. Typically, this involves scoring a conservative substitution as a partial rather than a full mismatch, thereby increasing the percentage sequence identity. Thus, for example, where an identical amino acid is given a score of 1 and a non-conservative substitution is given a score of zero, a conservative substitution is given a score between zero and 1. The scoring of conservative substitutions is calculated, e.g., as implemented in the program PC/GENE (Intelligenetics, Mountain View, Calif.).

As used herein, “percent sequence identity” means the value determined by comparing two optimally aligned sequences over a comparison window, wherein the portion of the polynucleotide sequence in the comparison window may comprise additions or

deletions (i.e., gaps) as compared to the reference sequence (which does not comprise additions or deletions) for optimal alignment of the two sequences. The percentage is calculated by determining the number of positions at which the identical nucleic acid base or amino acid residue occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the window of comparison, and multiplying the result by 100 to yield the percentage of sequence identity.

The term "substantial identity" of polynucleotide sequences means that a polynucleotide comprises a sequence that has at least 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, or 94%, or even at least 95%, 96%, 97%, 98%, 99% or 100% sequence identity, compared to a reference sequence using one of the alignment programs described herein using standard parameters. One of skill in the art will recognize that these values can be appropriately adjusted to determine corresponding identity of proteins encoded by two nucleotide sequences by taking into account codon degeneracy, amino acid similarity, reading frame positioning, and the like. Substantial identity of amino acid sequences for these purposes normally means sequence identity of at least 70% (e.g., 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%), at least 80% (e.g., 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%), at least 90% (e.g., 91%, 92%, 93%, or 94%), or even at least 95% (e.g., 96%, 97%, 98%, 99%, or 100%).

The term "substantial identity" in the context of a peptide indicates that a peptide comprises a sequence with at least 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, or 94%, or even 95%, 96%, 97%, 98% or 99%, sequence identity to the reference sequence over a specified comparison window. In certain embodiments, optimal alignment is conducted using the homology alignment algorithm of Needleman and Wunsch (Needleman and Wunsch, J. Mol. Biol., 48, 443 (1970)). An indication that two peptide sequences are substantially identical is that one peptide is immunologically reactive with antibodies raised against the second peptide. Thus, a peptide is substantially identical to a second peptide, for example, where the two peptides differ only by a conservative substitution. Thus, the disclosure also provides nucleic acid molecules and peptides that are substantially identical to the nucleic acid molecules and peptides presented herein.

Another indication that nucleotide sequences are substantially identical is if two molecules hybridize to each other under stringent conditions. Hybridization of nucleic acids is discussed in more detail below.

5 ***Oligonucleotide Primers and Probes***

As described herein, the methods provided in this disclosure rely upon oligonucleotides, sometimes referred to as primers or probes, to identify or detect features contained within the nucleic acid obtained from an individual, subject, or patient. The term “nucleic acid probe” or a “probe specific for” a nucleic acid refers to a nucleic acid sequence
10 that has at least about 80%, e.g., at least about 90%, e.g., at least about 95% contiguous sequence identity or homology to the nucleic acid sequence encoding the targeted sequence of interest. A probe (or oligonucleotide or primer) of the disclosure is at least about 8 nucleotides in length (e.g., at least about 8-50 nucleotides in length, e.g., at least about 10-40, e.g., at least about 15-35 nucleotides in length). The oligonucleotide probes or primers of the
15 disclosure may comprise at least about eight nucleotides at the 3' of the oligonucleotide that have at least about 80%, e.g., at least about 85%, e.g., at least about 90%, e.g., at least about 95% contiguous identity to the targeted sequence of interest.

Primer pairs are useful for determination of the nucleotide sequence of a particular SNP using PCR. The pairs of single-stranded DNA primers can be annealed to sequences
20 within or surrounding the SNP in order to prime amplifying DNA synthesis of the SNP itself. The first step of the process involves contacting a biological sample obtained from a subject, which sample contains nucleic acid, with at least one primer to form a hybridized DNA. The oligonucleotide primers that are useful in the methods of the present disclosure can be any primer comprised of about 8 bases up to about 80 or 100 bases or more. In one embodiment
25 of the present disclosure, the primers are between about 10 and about 20 bases.

The primers themselves can be synthesized using techniques that are well known in the art. Generally, the primers can be made using oligonucleotide synthesizing machines that are commercially available.

The primers or probes of the present disclosure can be labeled using techniques
30 known to those of skill in the art. For example, the labels used in the assays of disclosure can be primary labels (where the label comprises an element that is detected directly) or

secondary labels (where the detected label binds to a primary label, e.g., as is common in immunological labeling). An introduction to labels (also called "tags"), tagging or labeling procedures, and detection of labels is found in Polak and Van Noorden (1997) Introduction to Immunocytochemistry, second edition, Springer Verlag, N.Y. and in Haugland (1996)

5 Handbook of Fluorescent Probes and Research Chemicals, a combined handbook and catalogue Published by Molecular Probes, Inc., Eugene, Oreg. Primary and secondary labels can include undetected elements as well as detected elements. Useful primary and secondary labels in the present disclosure can include spectral labels such as fluorescent dyes (e.g., fluorescein and derivatives such as fluorescein isothiocyanate (FITC) and Oregon Green™, 10 rhodamine and derivatives (e.g., Texas red, tetramethylrhodamine isothiocyanate (TRITC), etc.), digoxigenin, biotin, phycoerythrin, AMCA, CyDyes™, and the like), radiolabels (e.g., ³H, ¹²⁵I, ³⁵S, ¹⁴C, ³²P, ³³P), enzymes (e.g., horse-radish peroxidase, alkaline phosphatase) spectral colorimetric labels such as colloidal gold or colored glass or plastic (e.g., polystyrene, polypropylene, latex) beads. The label may be coupled directly or indirectly to a 15 component of the detection assay (e.g., the labeled nucleic acid) according to methods well known in the art. As indicated above, a wide variety of labels may be used, with the choice of label depending on sensitivity required, ease of conjugation with the compound, stability requirements, available instrumentation, and disposal provisions.

In general, a detector that monitors a probe-substrate nucleic acid hybridization is 20 adapted to the particular label that is used. Typical detectors include spectrophotometers, phototubes and photodiodes, microscopes, scintillation counters, cameras, film and the like, as well as combinations thereof. Examples of suitable detectors are widely available from a variety of commercial sources known to persons of skill. Commonly, an optical image of a substrate comprising bound labeled nucleic acids is digitized for subsequent computer 25 analysis.

Labels include those that use (1) chemiluminescence (using Horseradish Peroxidase and/or Alkaline Phosphatase with substrates that produce photons as breakdown products) with kits being available, e.g., from Molecular Probes, Amersham, Boehringer-Mannheim, and Life Technologies/Gibco BRL; (2) color production (using both Horseradish Peroxidase 30 and/or Alkaline Phosphatase with substrates that produce a colored precipitate) (kits available from Life Technologies/Gibco BRL, and Boehringer-Mannheim); (3)

hemifluorescence using, e.g., Alkaline Phosphatase and the substrate AttoPhos (Amersham) or other substrates that produce fluorescent products, (4) fluorescence (e.g., using Cy-5 (Amersham), fluorescein, and other fluorescent labels); (5) radioactivity using kinase enzymes or other end-labeling approaches, nick translation, random priming, or PCR to incorporate radioactive molecules into the labeled nucleic acid. Other methods for labeling and detection will be readily apparent to one skilled in the art.

Fluorescent labels can be used and have the advantage of requiring fewer precautions in handling and being amenable to high-throughput visualization techniques (optical analysis including digitization of the image for analysis in an integrated system comprising a computer). Preferred labels are typically characterized by one or more of the following: high sensitivity, high stability, low background, low environmental sensitivity and high specificity in labeling. Fluorescent moieties, which can be incorporated into a label, generally are known including Texas red, dioxegenin, biotin, 1- and 2-aminonaphthalene, p,p'-diaminostilbenes, pyrenes, quaternary phenanthridine salts, 9-aminoacridines, p,p'-diaminobenzophenone imines, anthracenes, oxacarbocyanine, merocyanine, 3-aminoequilenin, perylene, bis-benzoxazole, bis-p-oxazolyl benzene, 1,2-benzophenazin, retinol, bis-3-aminopyridinium salts, hellebrigenin, tetracycline, sterophenol, benzimidazolylphenylamine, 2-oxo-3-chromen, indole, xanthen, 7-hydroxycoumarin, phenoxazine, calicylate, strophanthidin, porphyrins, triarylmethanes, flavin and many others. Many fluorescent labels are commercially available from the SIGMA Chemical Company (Saint Louis, MO), Molecular Probes, R&D systems (Minneapolis, MN), Pharmacia LKB Biotechnology (Piscataway, NJ), CLONTECH Laboratories, Inc. (Palo Alto, CA), Chem Genes Corp., Aldrich Chemical Company (Milwaukee, WI), Glen Research, Inc., GIBCO BRL Life Technologies, Inc. (Gaithersburg, MD), Fluka ChemicaBiochemika Analytika (Fluka Chemie AG, Buchs, Switzerland), and Applied Biosystems™ (Foster City, CA), as well as many other commercial sources known to one of skill.

Means of detecting and quantifying labels are well known to those of skill in the art. Thus, for example, when the label is a radioactive label, means for detection include a scintillation counter or photographic film as in autoradiography; and when the label is optically detectable, typical detectors include microscopes, cameras, phototubes, photodiodes, and many other detection systems that are widely available.

Oligonucleotide primers or probes may be prepared having any of a wide variety of base sequences according to techniques that are well known in the art. Suitable bases for preparing an oligonucleotide primer or probe may be selected from naturally occurring nucleotide bases such as adenine, cytosine, guanine, uracil, and thymine; and non-naturally occurring or "synthetic" nucleotide bases such as 7-deaza-guanine 8-oxo-guanine, 6-mercaptoguanine, 4-acetylcytidine, 5-(carboxyhydroxyethyl)uridine, 2'-O-methylcytidine, 5-carboxymethylamino-methyl-2-thioridine, 5-carboxymethylaminomethyluridine, dihydrouridine, 2'-O-methylpseudouridine, β ,D-galactosylqueosine, 2'-O-methylguanosine, inosine, N6-isopentenyladenosine, 1-methyladenosine, 1-methylpseudouridine, 1-methylguanosine, 1-methylinosine, 2,2-dimethylguanosine, 2-methyladenosine, 2-methylguanosine, 3-methylcytidine, 5-methylcytidine, N6-methyladenosine, 7-methylguanosine, 5-methylaminomethyluridine, 5-methoxyaminomethyl-2-thiouridine, β ,D-mannosylqueosine, 5-methoxycarbonylmethyluridine, 5-methoxyuridine, 2-methylthio-N6-isopentenyladenosine, N-((9- β -D-ribofuranosyl-2-methylthiopurine-6-yl)carbamoyl)threonine, N-((9- β -D-ribofuranosylpurine-6-yl)N-methyl-carbamoyl)threonine, uridine-5-oxyacetic acid methylester, uridine-5-oxyacetic acid, wybutosine, pseudouridine, queosine, 2-thiocytidine, 5-methyl-2-thiouridine, 2-thiouridine, 2-thiouridine, 5-Methyluridine, N-((9-beta-D-ribofuranosylpurine-6-yl)carbamoyl)threonine, 2'-O-methyl-5-methyluridine, 2'-O-methyluridine, wybutosine, and 3-(3-amino-3-carboxypropyl)uridine.

Any oligonucleotide backbone may be employed, including DNA, RNA (although RNA is less preferred than DNA), modified sugars such as carbocycles, and sugars containing 2' substitutions such as fluoro and methoxy. The oligonucleotides may be oligonucleotides wherein at least one, or all, of the internucleotide bridging phosphate residues are modified phosphates, such as methyl phosphonates, methyl phosphonotioates, phosphorinopholidates, phosphoropiperazidates and phosphoriamidates (for example, every other one of the internucleotide bridging phosphate residues may be modified as described). The oligonucleotide may be a "peptide nucleic acid" such as described in Nielsen *et al.*, Science, 254:1497-1500 (1991).

As used herein, a "single base pair extension probe" is a nucleic acid that selectively recognizes a single nucleotide polymorphism (i.e., either the A or the G of an A/G polymorphism). Generally, these probes take the form of a DNA primer (e.g., as in PCR

primers) that are modified so that incorporation of the primer releases a fluorophore. One example of this is a Taqman[®] probe that uses the 5' exonuclease activity of the enzyme Taq Polymerase for measuring the amount of target sequences in the samples. TaqMan[®] probes consist of a 18-22 bp oligonucleotide probe, which is labeled with a reporter fluorophore at the 5' end, and a quencher fluorophore at the 3' end. Incorporation of the probe molecule into a PCR chain (which occurs because the probe set is contained in a mixture of PCR primers) liberates the reporter fluorophore from the effects of the quencher. The primer must be able to recognize the target binding site. Some primer extension probes can be "activated" directly by DNA polymerase without a full PCR extension cycle.

The only requirement is that the oligonucleotide probe should possess a sequence at least a portion of which is capable of binding to a known portion of the sequence of the DNA sample. The nucleic acid probes provided by the present disclosure are useful for a number of purposes.

Methods of Detecting Nucleic Acids

A. Amplification

According to the methods of the present disclosure, the amplification of DNA present in a biological sample may be carried out by any means known to the art. Examples of suitable amplification techniques include, but are not limited to, polymerase chain reaction (including, for RNA amplification, reverse-transcriptase polymerase chain reaction), ligase chain reaction, strand displacement amplification, transcription-based amplification, self-sustained sequence replication (or "3SR"), the Qbeta replicase system, nucleic acid sequence-based amplification (or "NASBA"), the repair chain reaction (or "RCR"), and boomerang DNA amplification (or "BDA").

The bases incorporated into the amplification product can be natural or modified bases (modified before or after amplification), and the bases can be selected to optimize subsequent detection steps (e.g., electrochemical detection steps).

Polymerase chain reaction (PCR) can be carried out in accordance with known techniques. See, e.g., U.S. Patent Numbers 4,683,195; 4,683,202; 4,800,159; and 4,965,188. In general, PCR involves, first, treating a nucleic acid sample (e.g., in the presence of a heat stable DNA polymerase) with one oligonucleotide primer for each strand of the specific

sequence to be detected under hybridizing conditions so that an extension product of each primer is synthesized that is complementary to each nucleic acid strand, with the primers sufficiently complementary to each strand of the specific sequence to hybridize therewith so that the extension product synthesized from each primer, when it is separated from its complement, can serve as a template for synthesis of the extension product of the other primer, and then treating the sample under denaturing conditions to separate the primer extension products from their templates if the sequence or sequences to be detected are present. These steps are cyclically repeated until the desired degree of amplification is obtained. Detection of the amplified sequence may be carried out by adding, to the reaction product, an oligonucleotide probe capable of hybridizing to the reaction product (e.g., an oligonucleotide primer or probe of the present disclosure), the probe carrying a detectable label, and then detecting the label in accordance with known techniques. Various labels that can be incorporated into or operably linked to nucleic acids are well known in the art, such as radioactive, enzymatic, and fluorescent labels. Where the nucleic acid to be amplified is RNA, amplification may be carried out by initial conversion to DNA by reverse transcriptase in accordance with known techniques.

Strand displacement amplification (SDA) can be carried out in accordance with known techniques. For example, SDA can be carried out with a single amplification primer or a pair of amplification primers, with exponential amplification being achieved with the latter. In general, SDA amplification primers comprise, in the 5' to 3' direction, a flanking sequence (the DNA sequence of which is noncritical), a restriction site for the restriction enzyme employed in the reaction, and an oligonucleotide sequence (e.g., an oligonucleotide primer or probe as described herein) that hybridizes to the target sequence to be amplified and/or detected. The flanking sequence, which serves to facilitate binding of the restriction enzyme to the recognition site and provides a DNA polymerase priming site after the restriction site has been nicked, can be about 15 to 20 nucleotides in length. The restriction site is functional in the SDA reaction. For example, the oligonucleotide primer or probe portion can be about 13 to 15 nucleotides in length.

Ligase chain reaction (LCR) also can be carried out in accordance with known techniques. In general, the reaction is carried out with two pairs of oligonucleotide probes: one pair binds to one strand of the sequence to be detected; the other pair binds to the other

strand of the sequence to be detected. Each pair together completely overlaps the strand to which it corresponds. The reaction is carried out by, first, denaturing (e.g., separating) the strands of the sequence to be detected, then reacting the strands with the two pairs of oligonucleotide probes in the presence of a heat stable ligase so that each pair of oligonucleotide probes is ligated together, then separating the reaction product, and then cyclically repeating the process until the sequence has been amplified to the desired degree. Detection then can be carried out in like manner as described above with respect to PCR.

According to the methods described herein, a particular SNP at a particular locus can be detected. Techniques that are useful in the methods described herein include, but are not limited to, direct DNA sequencing, PFGE analysis, allele-specific oligonucleotide (ASO), dot blot analysis and denaturing gradient gel electrophoresis, and are well known to a skilled artisan.

There are several methods that can be used to detect DNA sequence variation. Direct DNA sequencing, either manual sequencing or automated fluorescent sequencing can detect sequence variation. Another approach is the single-stranded conformation polymorphism assay (SSCA). This method does not detect all sequence changes, especially if the DNA fragment size is greater than 200 bp but can be optimized to detect most DNA sequence variation. The reduced detection sensitivity is a disadvantage, but the increased throughput possible with SSCA makes it an attractive, viable alternative to direct sequencing for mutation detection on a research basis. The fragments that have shifted mobility on SSCA gels then can be sequenced to determine the exact nature of the DNA sequence variation. Other approaches based on the detection of mismatches between the two complementary DNA strands include clamped denaturing gel electrophoresis (CDGE), heteroduplex analysis (HA) and chemical mismatch cleavage (CMC). Once a sequence change has been identified, an allele specific detection approach such as allele specific oligonucleotide (ASO) hybridization can be utilized to rapidly screen large numbers of other samples for that same sequence change (e.g., mutation, polymorphism). Such a technique can utilize probes that are labeled with gold nanoparticles to yield a visual color result.

Detection of SNPs can be accomplished by sequencing the desired target region using techniques well known in the art. Alternatively, sequences can be amplified directly from a

genomic DNA preparation from subject tissue using known techniques. The DNA sequence of the amplified sequences then can be determined.

There are several well-known methods for a more complete, yet still indirect, test for confirming the presence of a mutant allele: 1) single stranded conformation analysis (SSCA); 2) denaturing gradient gel electrophoresis (DGGE); 3) RNase protection assays; 4) allele-specific oligonucleotides (ASOs); 5) the use of proteins which recognize nucleotide mismatches, such as the *E. coli* mutS protein; and/or 6) allele-specific PCR. For allele-specific PCR, primers are used that hybridize at their 3' ends to a particular allele. If the particular mutation is not present, an amplification product is not observed. Amplification Refractory Mutation System (ARMS) can also be used. Insertions and deletions of genes can also be detected by cloning, sequencing, and amplification. In addition, restriction fragment length polymorphism (RFLP) probes for the gene or surrounding marker genes can be used to score alteration of an allele or an insertion in a polymorphic fragment. Other techniques for detecting insertions and deletions as known in the art can be used.

In the first three methods (SSCA, DGGE and RNase protection assay), a new electrophoretic band appears. SSCA detects a band that migrates differentially because the sequence change causes a difference in single-strand, intramolecular base pairing. RNase protection involves cleavage of the mutant polynucleotide into two or more smaller fragments. DGGE detects differences in migration rates of mutant sequences compared to wild-type sequences, using a denaturing gradient gel. In an allele-specific oligonucleotide assay, an oligonucleotide is designed which detects a specific sequence, and the assay is performed by detecting the presence or absence of a hybridization signal. In the mutS assay, the protein binds only to sequences that contain a nucleotide mismatch in a heteroduplex between mutant and wild-type sequences.

Mismatches, according to the present disclosure, are hybridized nucleic acid duplexes in which the two strands are not 100% complementary. Lack of total homology may be due to deletions, insertions, inversions, or substitutions. Mismatch detection can be used to detect point mutations in the gene or in its mRNA product. While these techniques are less sensitive than sequencing, they are simpler to perform on a large number of samples. An example of a mismatch cleavage technique is the RNase protection method. The riboprobe and either mRNA or DNA isolated from the tumor tissue are annealed (hybridized) together and

subsequently digested with the enzyme RNase A to detect some mismatches in a duplex RNA structure. If a mismatch is detected by RNase A, it cleaves at the site of the mismatch. Thus, when the annealed RNA preparation is separated on an electrophoretic gel matrix, if a mismatch has been detected and cleaved by RNase A, an RNA product will be seen which is smaller than the full-length duplex RNA for the riboprobe and the mRNA or DNA. The riboprobe need not be the full length of the mRNA or gene but can be a segment of either. If the riboprobe includes only a segment of the mRNA or gene, it will be desirable to use a number of these probes to screen the whole mRNA sequence for mismatches.

In similar fashion, DNA probes can be used to detect mismatches, through enzymatic or chemical cleavage. Alternatively, mismatches can be detected by shifts in the electrophoretic mobility of mismatched duplexes relative to matched duplexes. With either riboprobes or DNA probes, the cellular mRNA or DNA that might contain a mutation can be amplified using PCR before hybridization.

B. Sequencing

Due to its sensitivity and relative simplicity in terms of both workflow and technique, Sanger sequencing is used in a variety of applications from targeted sequencing to confirming variants identified using orthogonal methods. Sanger sequencing utilizes a chain-termination method to provide the identity and order of nucleotide bases in a given strand of DNA. This method makes use of chemical analogues of the four nucleotide bases (i.e., ddNTPs), which are missing the hydroxyl group required for extension of the polynucleotide chains that form the DNA molecule. By mixing radiolabeled, and, later, fluorescent labeled, ddNTPs with template DNA, strands of each possible length are produced when the ddNTPs get randomly incorporated, terminating the chain. In contrast to Sanger sequencing, the Maxam and Gilbert used a chemical cleavage technique. The chief advantages of the Maxam-Gilbert technique compared with Sanger's method are that sequencing could be done from the original DNA fragment, instead of from enzymatic copies, no PCR is required, and this method is less susceptible to mistakes of secondary structures or enzymatic mistakes. The products generated in sequencing reactions can be resolved by ionophoresis on acrylamide gels or using capillary electrophoresis.

Another sequencing technique referred to as pyrosequencing was developed that uses a two-enzyme process in which adenosine triphosphate (ATP) sulfurylase is used to convert

pyrophosphate into ATP, which is then used as the substrate for luciferase, thus producing light proportional to the amount of pyrophosphate. Additional approaches for sequencing nucleic acids include emulsion polymerase chain reaction (PCR), reversible terminator, and sequencing by oligonucleotide ligation and detection. Capillary electrophoresis (CE) instruments also can be used for sequencing.

High-throughput sequencing techniques also have been developed, termed next-generation sequencing (NGS). NGS is massively parallel, sequencing millions of fragments simultaneously per run. This high-throughput process translates into sequencing hundreds to thousands of genes at one time. NGS also offers greater discovery power to detect novel or rare variants with deep sequencing. The spectrum of analysis of NGS can extend from a small number of genes to an entire genome. Whole-genome sequencing (WGS) and whole-exome sequencing (WES) provide the sequence of DNA bases across the genome and exome, respectively. Whole-transcriptome sequencing provides sequence information about coding and multiple noncoding forms of RNA to assess variations and gene expression levels across the entire transcriptome. Targeted sequencing covers a relatively small set of genes or targeted regions of interest (e.g., to determine the presence or absence of a SNP). Real-time sequencing and single-molecule sequencing (SMS), capable of accurately sequencing long strands of nucleic acid without an intermediary and without previous transcription or amplification also have been developed.

Nanopore-based sequencing technology detects the unique electrical signals of different molecules as they pass through the nanopore with a semiconductor-based electronic detection system. This technology makes for a high throughput, cost effective sequencing solution. At the heart of the technology is the biological nanopore, a protein pore embedded in a membrane, while the brains of the technology lie in the electronics of a semiconductor integrated circuit and proprietary chemistries. The electronic sensor technology embedded in the chip enables automatic membrane assembly and nanopore insertion, while allowing for active control of individual sensors on the circuit. See, e.g., Oxford Nanopore and Pacific Biosciences. Nanopore and electronic sensor sequencing technology can be used to directly determine DNA methylation using native, non-bisulfite DNA.

C. Hybridization

The phrase “hybridizing specifically to” refers to the binding, duplexing, or hybridizing of a molecule only to a particular nucleotide sequence under stringent conditions when that sequence is present in a complex mixture (e.g., total cellular) DNA or RNA.

- 5 “Bind(s) substantially” refers to complementary hybridization between a primer or probe nucleic acid and a target nucleic acid and embraces minor mismatches that can be accommodated by reducing the stringency of the hybridization media to achieve the desired detection of the target nucleic acid sequence.

Generally, stringent conditions are selected to be about 5°C lower than the thermal
10 melting point (T_m) for the specific sequence at a defined ionic strength and pH. However, stringent conditions encompass temperatures in the range of about 1°C to about 20°C, depending upon the desired degree of stringency as otherwise qualified herein. Nucleic acids that do not hybridize to each other under stringent conditions are still substantially identical if the polypeptides they encode are substantially identical. This may occur, e.g., when a copy of
15 a nucleic acid is created using the maximum codon degeneracy permitted by the genetic code. One indication that two nucleic acid sequences are substantially identical is when the polypeptide encoded by the first nucleic acid is immunologically cross reactive with the polypeptide encoded by the second nucleic acid.

“Stringent conditions” are those that (1) employ low ionic strength and high
20 temperature for washing, for example, 0.015 M NaCl / 0.0015 M sodium citrate (SSC); 0.1% sodium lauryl sulfate (SDS) at 50°C, or (2) employ a denaturing agent such as formamide during hybridization, e.g., 50% formamide with 0.1% bovine serum albumin / 0.1% Ficoll / 0.1% polyvinylpyrrolidone / 50 mM sodium phosphate buffer at pH 6.5 with 750 mM NaCl, 75 mM sodium citrate at 42°C. Another example is use of 50% formamide, 5 x SSC (0.75 M
25 NaCl, 0.075 M sodium citrate), 50 mM sodium phosphate (pH 6.8), 0.1% sodium pyrophosphate, 5x Denhardt's solution, sonicated salmon sperm DNA (50 µg/ml), 0.1% SDS, and 10% dextran sulfate at 42°C, with washes at 42°C in 0.2 x SSC and 0.1% SDS. Other examples of stringent conditions are well known in the art.

“Stringent hybridization conditions” and “stringent hybridization wash conditions” in
30 the context of nucleic acid hybridization experiments such as Southern and Northern hybridizations are sequence dependent and are different under different environmental

parameters. Longer sequences hybridize specifically at higher temperatures. The thermal melting point (T_m) is the temperature (under defined ionic strength and pH) at which 50% of the target sequence hybridizes to a perfectly matched primer or probe sequence. Specificity is typically the function of post-hybridization washes, the critical factors being the ionic strength and temperature of the final wash solution. For DNA-DNA hybrids, the T_m can be approximated from the equation of Meinkoth and Wahl (1984, *Anal. Biochem.*, 138(2):267-84); $T_m = 81.5^\circ\text{C} + 16.6 (\log M) + 0.41 (\%GC) - 0.61 (\% \text{ form}) - 500/L$; where M is the molarity of monovalent cations, %GC is the percentage of guanosine and cytosine nucleotides in the DNA, % form is the percentage of formamide in the hybridization solution, and L is the length of the hybrid in base pairs. T_m is reduced by about 1°C for each 1% of mismatching; thus, T_m , hybridization, and/or wash conditions can be adjusted to hybridize to sequences of the desired identity. For example, if sequences with >90% identity are sought, the T_m can be decreased 10°C . Generally, stringent conditions are selected to be about 5°C lower than the T_m for the specific sequence and its complement at a defined ionic strength and pH. However, severely stringent conditions can utilize a hybridization and/or wash at 1, 2, 3, or 4°C lower than the T_m ; moderately stringent conditions can utilize a hybridization and/or wash at 6, 7, 8, 9, or 10°C lower than the T_m ; low stringency conditions can utilize a hybridization and/or wash at 11, 12, 13, 14, 15, or 20°C lower than the T_m . Using the equation, hybridization and wash compositions, and desired temperature, those of ordinary skill will understand that variations in the stringency of hybridization and/or wash solutions are inherently described. If the desired degree of mismatching results in a temperature of less than 45°C (aqueous solution) or 32°C (formamide solution), the SSC concentration can be increased so that a higher temperature can be used. Generally, highly stringent hybridization and wash conditions are selected to be about 5°C lower than the T_m for the specific sequence at a defined ionic strength and pH.

An example of highly stringent wash conditions is 0.15 M NaCl at 72°C for about 15 minutes. An example of stringent wash conditions is a 0.2 x SSC wash at 65°C for 15 minutes. Often, a high stringency wash is preceded by a low stringency wash to remove background signal. An example of a medium stringency wash for a duplex of, e.g., more than 100 nucleotides, is 1 x SSC at 45°C for 15 minutes. For short nucleotide sequences (e.g., about 10 to 50 nucleotides), stringent conditions typically involve salt concentrations of less

than about 1.5 M, less than about 0.01 to 1.0 M, Na ion concentration (or other salts) at pH 7.0 to 8.3, and the temperature is typically at least about 30°C and at least about 60°C for long oligonucleotides (e.g., >50 nucleotides). Stringent conditions also can be achieved by the addition of destabilizing agents such as formamide. In general, a signal to noise ratio of 2x (or higher) than that observed for an unrelated oligonucleotide in the particular hybridization assay indicates detection of a specific hybridization. Nucleic acids that do not hybridize to each other under stringent conditions are still substantially identical if the proteins that they encode are substantially identical. This can occur, e.g., when a copy of a nucleic acid is created using the maximum codon degeneracy permitted by the genetic code.

Very stringent conditions can be equal to the T_m for a particular oligonucleotide. An example of stringent conditions for hybridization of complementary nucleic acids that have more than 100 complementary residues on a filter in a Southern or Northern blot is 50% formamide, e.g., hybridization in 50% formamide, 1 M NaCl, 1% SDS at 37°C, and a wash in 0.1 x SSC at 60 to 65°C. Exemplary low stringency conditions include hybridization with a buffer solution of 30 to 35% formamide, 1 M NaCl, 1% SDS (sodium dodecyl sulphate) at 37°C, and a wash in 1x to 2x SSC (20x SSC = 3.0 M NaCl/0.3 M trisodium citrate) at 50 to 55°C. Exemplary moderate stringency conditions include hybridization in 40 to 45% formamide, 1.0 M NaCl, 1% SDS at 37°C, and a wash in 0.5x to 1x SSC at 55 to 60°C.

“Northern analysis” or “Northern blotting” is a method used to identify RNA sequences that hybridize to a known probe such as an oligonucleotide, DNA fragment, cDNA or fragment thereof, or RNA fragment. The probe can be labeled with a radioisotope such as ^{32}P , by biotinylation or with an enzyme. The RNA to be analyzed can be usually electrophoretically separated on an agarose or polyacrylamide gel, transferred to nitrocellulose, nylon, or other suitable membrane, and hybridized with the probe, using standard techniques well known in the art.

Nucleic acid sample may be contacted with an oligonucleotide in any suitable manner known to those skilled in the art. For example, the DNA sample may be solubilized in solution, and contacted with the oligonucleotide by solubilizing the oligonucleotide in solution with the DNA sample under conditions that permit hybridization. Suitable conditions are well known to those skilled in the art. Alternatively, the DNA sample may be solubilized in solution with the oligonucleotide immobilized on a solid support, whereby the

DNA sample may be contacted with the oligonucleotide by immersing the solid support having the oligonucleotide immobilized thereon in the solution containing the DNA sample.

The term "substrate" refers to any solid support to which an oligonucleotide may be attached. The substrate material may be modified, covalently or otherwise, with coatings or functional groups to facilitate binding of oligonucleotides. Suitable substrate materials include polymers, glasses, semiconductors, papers, metals, gels and hydrogels among others. Substrates may have any physical shape or size, e.g., plates, strips, or microparticles. The term "spot" refers to a distinct location on a substrate to which oligonucleotides of known sequence are attached. A spot may be an area on a planar substrate, or it may be, for example, a microparticle distinguishable from other microparticles. The term "bound" means affixed to the solid substrate. A spot is "bound" to the solid substrate when it is affixed in a particular location on the substrate for purposes of the screening assay.

In certain embodiments of the present disclosure, the substrate is a polymer, glass, semiconductor, paper, metal, gel or hydrogel. In certain embodiments of the present disclosure, a kit can further include a solid substrate and at least one control oligonucleotide, wherein the at least one control oligonucleotide is bound onto the substrate in a distinct spot.

In certain embodiments of the present disclosure, the solid substrate is a microarray. An "array" or "microarray" is used synonymously herein to refer to a plurality of primers or probes attached to one or more distinguishable spots on a substrate. A microarray may include a single substrate or a plurality of substrates, for example a plurality of beads or microspheres. A "copy" of a microarray contains the same types and arrangements of primer or probes.

Methods for Detecting or Predicting Cardiovascular Disease or Estimating Survival

Better risk assessment for, or earlier detection of, cardiovascular disease is the first step toward more effective prevention. Those identified as being at higher risk (e.g., PPV of 69% for CHD) for CHD, CHD events, or as having CHD can be followed up promptly for further testing such as with coronary calcium or angiography, and more aggressive interventions. They can be tested or re-tested periodically for determining severity, identifying, customizing, optimizing intervention(s) (e.g., lifestyle, medical, therapeutics), management and monitoring. Conversely, those at lower risk (e.g., NPV of 99% for CHD) or

those free of CHD can be re-tested periodically to determine severity, identify, customize, optimize intervention(s) (e.g., lifestyle, medical, therapeutics), managed and monitored to ensure continued prevention due to the dynamic nature of DNA methylation. In addition, those with CVD (e.g., CHD) can be evaluated and their survivability estimated based on their genetic and/or epigenetic profile. It would be appreciated that, under some circumstances (e.g., for monitoring or follow-up purposes; for evaluating survivability of individuals at risk of or already determined to have CVD), the steps involved in detecting one or more SNPs may not be required, as that information may already be available (e.g., from a previous determination) or may not be necessary to predict CVD or survivability from CVD or severity or to identify, customize and optimize intervention(s) or to manage an individual.

Compared to the integrated genetic-epigenetic model, overall, conventional risk factors-based calculators and/or other detection tests such as stress test were considerably less sensitive, less generalizable, and also depicted a gender gap in performance. In contrast, the integrated genetic-epigenetic model described herein has the ability to capture and better understand the complex nature of CVD via three angles, genetics (inherited risk that is static), DNA methylation (acquired risk that is dynamic) and the genetic confounding of methylation signatures (i.e., $G + M + G \times M$).

The present disclosure provides a method for determining whether a subject has a likelihood of having CVD by determining methylation status of a CpG dinucleotide repeat or CpG dinucleotide repeat motif region, where the methylation status of the CpG dinucleotide is associated with CVD. However, the same principals apply to the assessment of the prevalence and/or incidence of a number of different types of CVD including, without limitation, coronary heart disease (CHD) (e.g., obstructive CHD), stroke, arrhythmia, cardiac arrest, congestive heart failure, atherosclerotic cardiovascular disease (ASCVD) and its associated cardiovascular events (CVE) including, for example, obstructive coronary artery disease (CAD), ischemia with no obstructive coronary arteries (INOCA), myocardial infarction (MI), stroke (e.g., TIA, hemorrhagic), and cardiovascular death. The present disclosure also provides methods for determining severity and/or estimating the survival of a subject having or at risk of having CVD. In certain embodiments, the method determines the methylation status of a plurality (e.g., any integer between 1 and 10,000, such as at least 100) of CpG dinucleotides and/or SNPs.

As used herein, a “biological sample” encompasses essentially any sample type obtained from a subject that can be used in a method as described herein. The biological sample may be any bodily fluid, tissue or any other sample from which clinically relevant biomarker levels may be determined. “Biological samples” also can encompass cells in culture, cell supernatants, cell lysates, blood, serum, plasma, urine, cerebral spinal fluid, biological fluid, and tissue samples. Various techniques and reagents find use in the methods of the present disclosure. In one embodiment of the disclosure, blood samples, or samples derived from blood, e.g., plasma, circulating, peripheral, lymphocytes, etc., are assayed for the presence of one or more SNPs and/or the methylation status of one or more CpG dinucleotides. A biological sample also can be saliva. Typically, a biological sample that contains nucleic acids is provided and tested. Biological samples can be derived from subjects using well known techniques such as finger prick, venipuncture, lumbar puncture, fluid sample such as saliva or urine, or tissue biopsy and the like.

As used herein, the term “healthy” means that a subject does not manifest a particular condition and is no more likely than at random to be susceptible to a particular condition.

Prevalence is defined by the American Psychological Association (APA) as the “the total number or percentage of cases (e.g., of a disease or disorder) existing in a population” (APA Dictionary of Psychology, (American Psychological Association, Washington, DC, 2007)). In some instances, point prevalence is used to describe the prevalence of cases at a discrete point of time, and period prevalence is used to describe the number of cases that exist for a period of time (e.g., a month, a year). Prevalence typically is expressed as a rate per population unit (e.g., number of cases per 100,000 people) instead of an absolute number or a percent.

Similarly, incidence is defined by the APA as “the rate of occurrence of new cases of a given event or condition (e.g., a disorder, disease, symptom, or injury) in a particular population in a given period” of time (APA Dictionary of Psychology, (American Psychological Association, Washington, DC, 2007)). As used herein, the term “incidence” is defined as a tendency or susceptibility for a subject to manifest a condition, in this case, CVD (e.g., CHD). In some instances, the period of time can be a year or less than a year; in some instances, the period of time can be longer than a year (e.g., two years, five years, ten years).

Diagnosis is defined by the APA as the “process of identifying and determining the nature of a disease or disorder by its signs and symptoms, through the use of assessment techniques (e.g., tests and examinations) and other available evidence” (APA Dictionary of Psychology, (American Psychological Association, Washington, DC, 2007)). A diagnosis
5 can refer to the present time period, or to a time period in the past or the future.

Likewise, prognosis is defined by the APA as “a prediction of the course, duration, severity, and outcome of a condition, disease, or disorder” (APA Dictionary of Psychology, (American Psychological Association, Washington, DC, 2007)). A prognosis can be made, for example, over a period of one month, six months, one year, five years, ten years, or
10 longer.

Risk assessment is defined as a “study of a subject done for the purpose of trying to determine the probability that that person will develop a particular disease or, if the disease is already present, the probability that the person will suffer exacerbation of it or death from it” (Youngson, 2005, Collins Dictionary of Medicine). In some instances, risk assessment is
15 based on conditions or events and not on disease. In some instances, a risk assessment is determined over a period of time (e.g., months, years).

Survival or survivability, as used herein, typically refers to the number of days an individual has remaining to live, measured from the point in time the biological sample (e.g., a saliva sample, a blood draw, etc.) was obtained and used to make this determination.

It would be understood by a skilled artisan that, in certain instances, “detection” of, e.g., a biomarker (e.g., the presence or absence of a biomarker) can result in a “diagnosis.” Similarly, it would be understood by a skilled artisan that, in certain instances, “detection” and “diagnosis” are used interchangeably.

Biomarkers are described herein that can be used in methods (e.g., predictive or
25 prognostic) of detecting CVD in a subject or estimating survival in a subject having CVD or at risk of having CVD. Such methods typically include providing a biological sample from the subject; contacting DNA from the biological sample with bisulfite under alkaline conditions; contacting the bisulfite-treated DNA with at least one first oligonucleotide primer at least 8 nucleotides in length that is complementary to a sequence that comprises a CpG
30 dinucleotide (e.g., at a GC locus referred to as cg04988978, cg21161138, cg12655112, cg03725309, cg12586707, and cg17901584, or another biomarker from Appendix A); and

determining the methylation status of the CpG dinucleotide. It would be understood that the at least one first oligonucleotide probe can detect either the unmethylated CpG dinucleotide or the methylated CpG dinucleotide. Such a method can further include determining the genotype of a single nucleotide polymorphism (SNP) (e.g., rs2869675, rs4376434, rs12129789, rs7585056, rs710987, rs4639796, rs1333048, rs12714414, rs942317, and rs1441433, or another biomarker from Appendix C) or a second SNP in linkage disequilibrium with the first SNP. As described herein, methylation of one or more particular CpG dinucleotides and the presence of one or more particular SNPs can be used to predict CVD in the subject. Also as described herein, methylation of one or more particular CpG dinucleotides and/or the presence of one or more particular SNPs can be used to estimate survivability of CVD.

In some embodiments, the method further comprises contacting the bisulfite-treated DNA with at least one second oligonucleotide probe at least 8 nucleotides in length that is complementary to a sequence that comprises a CpG dinucleotide, where the at least one second oligonucleotide probe detects either the unmethylated CpG dinucleotide or the methylated CpG dinucleotide, whichever is not detected by the at least one first oligonucleotide probe.

In some embodiments, the ratio of methylated CpG dinucleotides to unmethylated CpG dinucleotides in the biological sample can be determined as a part of the methods described herein. Determining the ratio of methylated CpG dinucleotides to unmethylated CpG dinucleotides can allow for a risk or outcome to be estimated or determined.

It would be appreciated that determining the methylation status of the one or more CpG dinucleotides and determining the presence (or absence) of a SNP can utilize any number of techniques, such as, for example, amplifying and/or sequencing steps. Amplifying and sequencing are well known techniques in the art and are used routinely to determine both the methylation status of a particular sequence and the presence / absence of a SNP.

Methods of determining the presence of biomarkers associated with CHD in a biological sample from a subject are provided. A similar approach can be used for any other form of CVD as well. Such methods typically include providing a first portion of the biological sample and contacting DNA from the first portion with bisulfite under alkaline conditions. The bisulfite-treated first portion can be contacted with a first oligonucleotide

probe that is at least 8 nucleotides in length and that is complementary to a sequence that comprises a CpG dinucleotide (detected, e.g., at a CG locus referred to as cg04988978, cg21161138, cg12655112, cg03725309, cg12586707, and cg17901584, or another biomarker from Appendix A), and, if necessary or desired, a second portion of the biological sample can
 5 be contacted with a nucleic acid probe at least 8 nucleotides in length that is complementary to a SNP (e.g., rs2869675, rs4376434, rs12129789, rs7585056, rs710987, rs4639796, rs1333048, rs12714414, rs942317, and rs1441433, or another biomarker from Appendix C).

As described herein, the percentage of methylation of the CpG dinucleotide at one or more of the GC loci designated cg04988978, cg21161138, cg12655112, cg03725309,
 10 cg12586707, and cg17901584 (or at a CpG dinucleotide that is in linkage disequilibrium with one or more of such CpG dinucleotides) and the identity of the nucleotide at one or more SNPs designated rs2869675, rs4376434, rs12129789, rs7585056, rs710987, rs4639796, rs1333048, rs12714414, rs942317, and rs1441433 (or at a SNP that is in linkage
 15 disequilibrium with one or more of such SNPs) are biomarkers associated with CVD and can be used to predict the likelihood that an individual will develop CVD and/or prognosticate as to the severity of the disease or the outcome (e.g., survival) for the individual.

In addition to the SNP and/or CpG biomarkers identified herein, one or more clinical indicators can be used to aid in either or both diagnostics and prognostics, interventions (e.g., lifestyle, therapeutic, medical) selection or customization or optimization, management, or
 20 monitoring. Without limitation, such clinical indicators can include demographics (e.g., age, sex, race); vital signs (e.g., heart rate (beats/min), systolic BP (mm Hg), diastolic BP (mm Hg)); medical history (e.g., smoking, atrial fibrillation/flutter, hypertension, coronary heart disease, myocardial infarction, heart failure, peripheral artery disease, COPD, diabetes (type 1 or type 2), CVA/TIA, chronic kidney disease, hemodialysis, angioplasty (peripheral or
 25 coronary), stent (peripheral or coronary), CABG, percutaneous coronary intervention); medications (ACE-I/ARB, beta blocker, aldosterone antagonist, loop diuretics, nitrates, CCB, statin, aspirin, warfarin, clopidogrel); coronary computed tomography angiography (e.g., aortic stenosis, FFR-CT, plaque type, total plaque); echocardiographic results (e.g., LVEF (%), RSVP (mm Hg)); stress test results (e.g., ischemia on scan, ischemia on ECG);
 30 angiography results (e.g., $\geq 70\%$ coronary stenosis in ≥ 2 vessels, $\geq 70\%$ coronary stenosis in ≥ 3 vessels); and/or lab measures (e.g., sodium, blood urea nitrogen (mg/dL), creatinine

(mg/dL), eGFR (median, CKDEPI), total cholesterol (mg/dL), LDL cholesterol (mg/dL), Ribitol, Hemoglobin, Hematocrit, Triglycerides, Alkaline Phosphatase, HbA1c, HDL-C, Non-HDL-C, ApoB, LDL-P¹, HDL-P¹, sdLDL-C, VLDL-C, Lp(a), hs-CRP, LpPLA₂ Activity, HOMOCYSTEINE, B TYPE NATRIURETIC PEPTIDE, glycohemoglobin (%),
 5 glucose (mg/dL), HGB (mg/dL), C-reactive protein (mg/L)), NT-proBNP, KIM-1, osteopontin, TIMP-1, kidney injury molecule-1, N-terminal pro B-type natriuretic peptide, osteopontin, tissue inhibitor of metalloproteinase-1, Uridine, Carotene-3, Ribitol, 1-stearoyl-2-adrenoyl-GPC, N-acetyl-isoputrescine, Lysophosphatidylcholine, Vanillactate acid, 3-ureidopropionate, Serum paraoxonase, Bone morphogenetic protein 1, Carboxypeptidase B2,
 10 Albumin, Histone H2B type 1-K, Versican core protein, Insulin-like growth factor-binding protein 2, Matrix-remodeling associated protein 5. etc.).

Kits for Detecting Cardiovascular Disease (CVD)

In a further embodiment of the disclosure, articles of manufacture and kits containing
 15 probes, oligonucleotides and/or antibodies are provided. Such articles of manufacture can be used in the methods described herein. An article of manufacture can include one or more containers with, for example, a label. Suitable containers include, for example, bottles, vials, and test tubes. The containers can be formed from a variety of materials such as glass or plastic. The container can hold a composition that includes one or more agents that are
 20 effective for practicing the methods described herein. The label on the container indicates that the composition can be used for a specific application. The kit of the disclosure will typically comprise the container described above and one or more other containers comprising materials desirable from a commercial and user standpoint, including buffers, diluents, filters, and package inserts with instructions for use.

25 In certain embodiments, the present disclosure provides a kit for determining the methylation status of at least one CpG dinucleotide and, in some cases, also the presence of at least one single-nucleotide polymorphism (SNP). In certain embodiments, a kit as described herein may contain a number of primers that is any integer between 1 and 10,000, such as 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, . . . 9997, 9998, 9999, 10,000. As used herein, the term
 30 "nucleic acid primer" or "nucleic acid probes" or "oligonucleotide" encompasses both DNA

and RNA sequences. In certain embodiments, the primers or probes may be physically located on a single solid substrate or on multiple substrates.

A kit as described herein can include at least one first nucleic acid primer (e.g., at least 8 nucleotides in length) that is complementary to a bisulfite-converted nucleic acid sequence comprising a CpG dinucleotide (detected, e.g., at a GC locus referred to as cg04988978, cg21161138, cg12655112, cg03725309, cg12586707, and cg17901584), and, in some instances, at least one second nucleic acid primer (e.g., at least 8 nucleotides in length) that is complementary to a SNP (e.g., rs2869675, rs4376434, rs12129789, rs7585056, rs710987, rs4639796, rs1333048, rs12714414, rs942317, and rs1441433). The at least one first nucleic acid primer can detect the methylated or unmethylated CpG dinucleotide.

It would be appreciated that any of the nucleic acid primers, probes or oligonucleotides described herein can include one or more nucleotide analogs and/or one or more synthetic or non-natural nucleotides.

It also would be appreciated that the kits described herein can include a solid substrate. In some embodiments, one or more of the nucleic acid primers can be bound to the solid support. Examples of solid supports include, without limitation, polymers, glass, semiconductors, papers, metals, gels, or hydrogels. Additional examples of solid supports include, without limitation, microarrays, or microfluidics cards.

It also would be appreciated that any of the kits described herein can include one or more detectable labels. In some embodiments, one or more of the nucleic acid primers can be labeled with the one or more detectable labels. Representative detectable labels include, without limitation, an enzyme label, a fluorescent label, and a colorimetric label.

Algorithm for Predicting Cardiovascular Disease (CVD) or Estimating Survivability from CVD

Any number of algorithms can be used including, without limitation, statistical algorithms (e.g., linear regression, logistic regression, proportional hazard models, etc.), machine learning algorithms (e.g., Random Forest, Gradient Boosting, Support Vector Machines, Neural Networks (e.g., deep neural network, extreme learning machine (ELM)), Bayes classifiers, Hidden Markov model, etc.), deep learning algorithms (e.g., convolutional neural networks, recurrent neural networks, autoencoders, large language model, etc.), time

series algorithms (e.g., ARIMA, etc.), Bayesian model algorithms (e.g., Bayesian Networks, etc.), and/or financial algorithms (e.g., decision tree, discrete event simulation, budget impact, etc.). See, for example, McKinney et al., 2011, *Appl. Bioinform.*, 5(2):77-88; Gunther et al., 2012, *BMC Genet.*, 13:37; and Ogutu et al., 2011, *BMC Proceedings*, 5(Suppl 3):S11. Any type of machine learning algorithm or deep learning neural network algorithm (tuned or non-tuned) capable of capturing linear and/or non-linear contribution of traits for the prediction can be used. In some instances, a combination of algorithms (e.g., a combination or ensemble of multiple algorithms that capture linear and/or non-linear contributions of traits) is used.

Furthermore, algorithm(s) can implement any one or more of: a regression algorithm, an instance-based method (e.g., k-nearest neighbor, learning vector quantization, self-organizing map, etc.), a regularization method, a decision tree learning method (e.g., classification and regression tree, chi-squared approach, Random Forest approach, multivariate adaptive approach, gradient boosting machine approach, etc.), a Bayesian method (e.g., naïve Bayes, Bayesian belief network, etc.), a kernel method (e.g., a support vector machine, a linear discriminant analysis, etc.), a clustering method (e.g., k-means clustering), an associated rule learning algorithm (e.g., an a priori algorithm), an artificial neural network model (e.g., a back-propagation method, a Hopfield network method, a learning vector quantization method, etc.), a deep learning algorithm (e.g., a Boltzmann machine, a convolution network method, a stacked auto-encoder method, etc.), a dimensionality reduction method (e.g., principal component analysis, partial least squares regression, etc.), an ensemble method (e.g., boosting, boot strapped aggregation, gradient boosting machine approach, etc.), and any other suitable algorithm.

Simply by way of example, Random Forest™ is a popular machine learning algorithm created by Breiman & Cutler for generating “classification trees” (see, for example, “stat.berkeley.edu/~breiman/RandomForests/cc_home.htm” on the World Wide Web). Using standard machine learning and predictive modeling techniques, a diagnostic classifier algorithm was written to be implemented in R and Python programming languages (though it can be implemented in many other programming languages), according to well described guidelines by Breiman & Cutler. A diagnostic classifier algorithm was generated using data from at least two traits (T) and the diagnosis of interest from that population. To

determine the output (e.g., diagnosis) for a new individual, one simply determines values for the at least two traits (T) and inputs that information into an algorithm (e.g., the diagnostic classifier algorithm described herein or another algorithm discussed above) that is capable of capturing the linear and non-linear contributions of the traits.

5 Prior to fitting model(s), input data can be conditioned or otherwise pre-processed, such that conditioned data elements (e.g., genomic reads associated with loci of interest, functional data, sensor data, other lifestyle data, etc.) are suitable for further processing. Conditioning, as described herein, can include filtering of data (e.g., sensor data outputs that have confidence values below a threshold etc.). Pre-processing step(s) can be taken on the
10 model input(s) such as dimensionality reduction (e.g., principal component analysis, linear discriminant analysis, auto encoders, uniform manifold approximation and projection, partial least squares regression, etc.) before training model(s).

 The model prediction(s) (e.g., risk score, diagnostics, etc.) uncertainty can be computed. The methods for estimating may include bootstrap methods, Bayesian methods,
15 Monte Carlo dropout, ensemble methods, sensitivity analysis etc. The uncertainty of the method and/or model described herein can be aggregated to provide overall system uncertainty. Uncertainty quantification can encompass all aspects of marker measurements. For instance, in epigenetic measurement, uncertainty quantification may include, but is not limited to, sampling error, reagent quality, instrumentation, and human error.

20 The uncertainty quantification mentioned herein, which can be used to perform quality control. For example, known reference sample(s) and/or measurement(s) can be compared against the new measurement(s). The difference between the known and new measurement(s), along with their uncertainty, can be used to evaluate error(s) and/or uncertainty(ies) and how they compare to a defined acceptable threshold. This process assists
25 in identifying sources of errors and/or uncertainties and minimizes such error(s) and/or uncertainty(ies) to ensure the reliability of the measurement(s). For example, it can be used to determine if a sample(s) requires re-measurement(s) or re-collection(s) to meet a defined acceptable threshold for the measurement(s) and/or marker(s).

 Returned classification regression and/or other outputs of model(s) can include
30 returned confidence-associated parameters in such classifications. In particular, confidence-associated parameters can have a score (e.g., percentile, other score) that indicates confidence

in the returned output. The confidence may be estimated by aggregating measurement and or modeling uncertainties. Transforming output data also can be used to enhance the expandability of the models described previously. For example, SHapley Additive
exPlanations (SHAP), Local Interpretable Model-agnostic Explanations, Integrated
5 Gradients, Partial Dependence Plots, Global Surrogate Models, etc., can be used to enhance
expandability of the output for users. One or more of these approaches can be utilized
concurrently. In specific examples, SHAP can be utilized to identify the most significant
contributing marker(s) to conditions or indication.

Additionally, or alternatively, dynamic aspects (e.g., changes over time in markers,
10 changes in frequency between instances of respective features, other temporal aspects, other
frequency-related aspects, etc.) of features derived from the samples can be used to predict or
otherwise anticipate health condition statuses to generate personalized intervention plans.

Samples can be collected once (e.g., at a single time point), or at a number of time
points (e.g., at random points, at regular points, in relation to triggering events, with other
15 frequency, etc.).

As described herein, the inputs can be at least one genotype (e.g., SNP) and/or the
methylation status of at least one CpG dinucleotide and/or other data, and the outcome can
represent a positive or a negative probability for CVD, however, severity also can be
evaluated. The Traits (T) used to determine the outcome can represent the methylation status
20 of at least one CpG dinucleotide or at least one genotype (e.g., of a SNP), but Traits (T) also
can correspond to at least one interaction (e.g., between methylation status and genotype
(CpGxSNP), between the methylation status of two different sites (CpGxCpG) or between
two different genotypes (SNPxSNP)). It would be appreciated that any such interactions can
be visualized using partial dependence plots.

25 The inputs also can be data or information (e.g., a dataset) including, without
limitation, clinical diagnoses; demographics (e.g., gender, race); lifestyle; imaging (e.g.,
cardiac CT scan, cardiac MRI, coronary angiogram); features derived from imaging (e.g.,
FFR from CT scan, percent stenosis); results from electrocardiogram or echocardiogram test;
results from stress tests; blood tests (e.g., for metabolic assays, genetic, epigenetic, protein,
30 etc.); blood pressure; results from a carotid ultrasound; and combinations thereof.

A dataset can include, without limitation, data derived from one or more of: body weight (e.g., receiving bodyweight values of the patients generated from a digital weighing scale), body fat percent, muscle mass, body water, height or other length measurements (e.g., via a ruler or measuring tape), other body mass index (BMI)- associated parameters, blood chemical and biochemical information, inflammatory markers, fasting blood sugar, high density lipids, low density lipids, blood interleukins, c-reactive protein, blood cell counts, electrophysiology signals (e.g., electroencephalogram signals, electromyography signals, galvanic skin response signals, electrocardiogram signals, etc.), heart rate, body temperature, cardiovascular parameters, continuous glucose monitoring (glycemic response), respiration parameters (e.g., respiration rate, depth/shalowness of breath, etc.), blood oxygenation signals, motion parameters, and any other suitable physiologically relevant parameter of the patient. Additionally, or alternatively, a dataset can include data derived from one or more of: electronic health records, health plan claims, questionnaires, survey, wearables, public sources (e.g., repositories) and any other direct or indirect data relevant to an individual. Additionally, or alternatively, a dataset can include data that is raw, imputed, transformed, longitudinal, cross-sectional or temporal.

FIG. 1 is a block diagram of an example CVD classification system 100. In some embodiments, the system 100 can perform monitoring and/or prediction of CVD. For example, the system 100 can be used to perform one or more of the example processes described herein.

In the illustrated example, a subject 101 provides a subject sample 102. In some embodiments, the subject sample 102 can be a blood sample, a saliva sample, a mucus sample, a urine or stool sample, or any other appropriate biological sample from the subject 101. In some embodiments, medical personnel 103 (e.g., a doctor, a nurse, a lab technician, a caregiver) may assist the subject 101 with obtaining the subject sample 102. In some embodiments, the subject 101 may obtain the subject sample 102 from herself or himself (e.g., by using a portable blood sampling device or a home collection kit).

A nucleic acid isolation module 110 isolates a nucleic acid sample 112 from the subject sample 102. In some embodiments, the nucleic acid isolation module 110 can be a manual, semi-automated, or automatic process that perform or more of cell lysis, removal of contaminating proteins, deactivating DNAases and/or RNAases, and recovery of DNA and/or

RNA. For example, the nucleic acid isolation module 110 can be a part of an automated process or analysis device configured to isolate the nucleic acid sample 112 from the subject sample 102. In another example, the nucleic acid isolation module 110 can be part of one or more of the example kits described in this document, to be used by a human user such as the medical personnel 103.

A genotyping assay module 120 receives a portion 114a of the nucleic acid sample 112. The genotyping assay module 120 is configured to perform a genotyping assay on the portion 114a of the nucleic acid sample 112 to detect the presence of at least one SNP, wherein the at least one SNP is a first SNP selected from rs2869675, rs4376434, rs12129789, rs7585056, rs710987, rs4639796, rs1333048, rs12714414, rs942317, and rs1441433 or from Appendix C and/or is a second SNP in linkage disequilibrium ($R > 0.3$) with a first SNP selected from rs2869675, rs4376434, rs12129789, rs7585056, rs710987, rs4639796, rs1333048, rs12714414, rs942317, and rs1441433 or from Appendix C to determine, identify, or otherwise obtain a collection of genotype data 122. In some embodiments, the genotyping assay module 120 can be a manual, semi-automated, or automatic process. For example, genotyping assay module 120 can be a part of an automated process or analysis device configured to perform a genotyping assay on the portion 114a. In another example, the genotyping assay module 120 can be part of one or more of the example kits described in this document, to be used by a human user such as the medical personnel 103 or a laboratory technician.

A methylation assay module 130 receives a portion 114b of the nucleic acid sample 112. The methylation assay module 130 is configured to bisulfite convert the nucleic acid in the portion 114b of the nucleic acid sample 112 and perform methylation assessment on the portion 114b of the nucleic acid sample 112 to detect methylation status of at least one CpG site selected from cg04988978, cg21161138, cg12655112, cg03725309, cg12586707, and cg17901584 or from Appendix A and/or a CpG site collinear ($R > 0.3$) with a CpG site selected from cg04988978, cg21161138, cg12655112, cg03725309, cg12586707, and cg17901584 or from Appendix A to determine, identify, or otherwise obtain a collection of methylation data 132.

An identification system 140 is configured to receive the collection of genotype data 122 and the collection of methylation data 132 and identify one or more predetermined traits

or characteristics of the subject 101 based on a diagnostic classifier algorithm module 142. The diagnostic classifier algorithm module 142 is 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. In some embodiments, the diagnostic classifier algorithm module 142 can perform one or more of the algorithms described herein that may indicate the presence of disease (e.g., diagnostic indicators) or a propensity to develop disease (e.g., predict) or the severity of disease or the selection, customization, or optimization of one or more intervention(s) (e.g., lifestyle, medical, therapeutic) or effectiveness of management or the monitoring of disease, severity or risk. For example, the identification system may be configured to identify genetic and/or environmental characteristics that determines the presence of or the likelihood of a subject developing disease (e.g., cardiovascular disease), even when the disease is of polygenic origin. In some implementations, the diagnostic classifier algorithm module 142 can be a machine learning algorithm capable of accounting for linear and non-linear effects.

The identification system 840 provides an output 150 based on the diagnostic and/or prognostic indicators provided by the diagnostic classifier algorithm module 142. In some embodiments, the identification system 140 can include an output module configured to provide the output 150. In some implementations, the output 150 can be an identification of one or more diseases that the subject 101 may already have. For example, the output 150 may indicate that traits that are indicative of the presence of cardiovascular disease were found in the subject 101. In some implementations, the output 150 can be an indication of a likelihood that the subject 101 may develop a disease within a predetermined time frame (e.g., the subject 101 may have a 43% chance of developing cardiovascular disease within 3 years, the subject 101 may have a 77% of having a heart attack within 2 years). In some implementations, the output 150 can include therapeutic and/or preventative recommendations based on the diagnostic and/or prognostic indicators provided by the diagnostic classifier algorithm module 142. For example, in response to an identification or prediction of a diabetic or cardiac condition in the subject 101, the output 150 may include a recommendation to consult with the medical personnel 103, identify possible dietary or lifestyle changes by the subject 101 to address or avoid the condition, identify potential interventions and/or remedies for the subject 101 to consider in consultation with the medical personnel 103, or combinations of these and/or any other appropriate information based on

the output of the algorithm(s) of the diagnostic classifier algorithm module 142. In some instances, the output 150 can be an estimate of survivability of a subject at risk for or that has been determined to have CVD.

In the illustrated example, the output 150 is provided in various formats. The
5 information provided by the output 150 can be formatted into a message 160 that is provided to the subject 101 and/or to the medical personnel 103. In some implementations, the message 160 can be formatted as a report (e.g., a word processing file, a portable document format file) that is at least temporarily stored on a non-transitory storage medium (e.g., a hard drive, a FLASH memory), where it can be retrieved by the subject 101 and/or the medical
10 personnel 103 for review. In some implementations, the message 160 can be formatted as an electronic message (e.g., an email, a text message, an instant message) that is transmitted to the subject 101 and/or the medical personnel 103 for review. In some implementations, the message 160 can be a printed report. For example, the output 150 can be provided to a printing system that is configured to generate a hard copy report based on the output 150.
15 Subsequent automated or manual processing systems can package the report as a letter or other parcel that can be sent for physical delivery to the subject 101 and/or to the medical personnel 103 (e.g., the system 100 can create a paper printout of the results and mail them through postal mail).

A treatment device 170 can be configured to receive the diagnostic and/or prognostic
20 indicators provided by the output 150 and provide interventions (e.g., lifestyle, therapeutic, medical) based on the diagnostic and/or prognostic indicators. For example, the output 150 may indicate that the subject 101 has a high likelihood of suffering cardiac arrest within the next two years, and the treatment device 170 may be a drug (e.g., a tablet or capsule) or an implantable drug delivery system that reacts by identifying or by receiving configuration
25 settings for an appropriate dosage of a statin, acetylsalicylic acid (aspirin), an anti-inflammatory drug, a blood thinner, or combinations of these and/or any other appropriate therapeutic and/or preventative substances. In some embodiments, the treatment device 170 can be configured to also include one or more of the nucleic acid isolation module 110, the genotyping assay module 120, the methylation assay module 130, or the identification system
30 140.

A storage system 180 is configured to store the output 150. For example, the information included in the output 150 can be stored temporarily, for a predetermined period of time, or substantially permanently in a database, in a file, or as any other appropriate collection of data. In some embodiments, the storage system 180 can store the output 150 in a non-transitory storage medium (e.g., a hard drive, a FLASH memory). For example, the output 150 may include some or all of the collection of genotype data 122, the collection of methylation data 132, and/or the output 150 in personal health record that the subject 101 can store or carry with them. In some embodiments, the storage system 180 can store the output 150 as a physical medium, for example, the storage system 180 can include a printer that can generate a paper report based on the output 150, and/or store the report as a hard copy that can be physically filed away for later retrieval.

An input/output device 182 is physical device configured to display or otherwise present an output that is perceptible to humans (e.g., the subject 101, the medical personnel 103). For example, the input/output device 182 may be an electronic display device in a doctor's office. The system 100 may process the subject sample 102, and then alter the configuration of pixels onscreen to modify the information displayed by the input/output device 182 based on the output 150 (e.g., a screen can be updated to display an identified diagnosis and/or prognosis for the subject 101 to the medical personnel). In another example, the input/output device 182 can be configured to provide audible (e.g., spoken output) and/or tactile (e.g., braille, haptic, vibratory) output that modifies or otherwise transforms the output 150 into a physical and/or tangible output (e.g., to convey the diagnostic and/or prognostic indicators in a manner that is perceptible to a user who is sight-challenged). In another example, the input/output device 182 can be configured to alter, transform, or modify a physical characteristic of a physical structure or medium based on the output 150.

A user device 184 (e.g., a computer, a smartphone, a tablet computer, a computerized terminal) is configured to display, emit, or otherwise present one or more outputs that are perceptible to a human user, such as the subject 101 and/or the medical personnel 103. For example, the user device 184 can receive the output 150 (e.g., as data, as the message 160) and provide an alert to the user and/or provide an output (e.g., display a report, read a report aloud) based on the output 150. In some embodiments, the user device 184 can include one or more of the storage device 180 or the input/output device 182. In some embodiments, the

user device 182 can be part of the treatment device 170. In some embodiments, the user device 184 can be configured to include one or more of the nucleic acid isolation module 110, the genotyping assay module 120, the methylation assay module 130, or the identification system 140.

5 In some implementations, some or all of the system 100 may be reused to provide additional information. For example, the system 100 may be used to gather an initial set of health information for the subject 101 and/or identify information that can assist the medical personnel 103 with an initial diagnosis/prognosis. Later, the patient 101 may be re-examined using the system 100, for example, to determine the effectiveness of prescribed medical
10 and/or lifestyle strategies over time. Since the collection of genotype data 122 does not change over time for an individual person, the system 100 may refrain from performing the functions of the genotyping assay module 120 again. In such examples, the methylation assay module 130 may be used to generate an updated version of the collection of methylation data 132, and the updated collection of methylation data 132 can be provided to the identification
15 system 140 for processing along with the collection of genotype data 122 that was previously generated. In some implementations, the subject sample 102 can be collected on a periodic basis and processed based on the existing collection of genotype data 122 and updated collections of methylation data 132 to produce updated outputs 150 that can be used to provide ongoing monitoring of one or more conditions identified for the subject 101.

20 FIG. 2 is a flow diagram of an example process 200 for cardiovascular disease classification. In some implementations, the process 200 can be some or all of the example processes described above. In some implementations, the process 200 can be the process performed by some or all of the example system 100 of FIG. 1.

25 At 210, a nucleic acid sample is isolated from a subject sample. For example, the example nucleic acid isolation module 110 can be configured to isolate and/or substantially purify nucleic acid compositions from the example subject sample 102 to produce the example nucleic acid sample 112.

30 At 220, a genotyping assay is performed 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 selected from rs2869675, rs4376434, rs12129789, rs7585056, rs710987, rs4639796, rs1333048, rs12714414, rs942317, and rs1441433 or from Appendix C and/or is a second

SNP in linkage disequilibrium ($R > 0.3$) with a first SNP selected from rs2869675, rs4376434, rs12129789, rs7585056, rs710987, rs4639796, rs1333048, rs12714414, rs942317, and rs1441433 or from Appendix C to obtain genotype data. For example, the example genotyping assay module 120 could be used to analyze the example portion 114a of the nucleic acid sample 112 to produce the example collection of genotype data 122.

At 230, a second portion of the nucleic acid sample is bisulfite converted, and a methylation assessment is performed on the second portion of the nucleic acid sample to detect methylation status of at least one CpG site selected from cg04988978, cg21161138, cg12655112, cg03725309, cg12586707, and cg17901584 or from Appendix A and/or a CpG site collinear ($R > 0.3$) with a CpG site selected from cg04988978, cg21161138, cg12655112, cg03725309, cg12586707, and cg17901584 or from Appendix A to obtain methylation data. For example, the example methylation assay module 130 can be used to process the portion 114b of the nucleic acid sample 112 to produce the example collection of methylation data 132.

At 240, the genotype data from step 220 and/or methylation data from step 230 is input into an algorithm. For example, the example collection of genotype data 122 and the example collection of methylation data 132 are input into the example identification system 140 and processed using the example diagnostic classifier algorithm module 142.

At 250, at least one SNP main effect and/or at least one CpG main effect and/or at least one interaction effect are accounted for. For example, the example diagnostic classifier algorithm module 142 can be 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. In some implementations, the diagnostic classifier algorithm module 142, can be a machine learning algorithm capable of accounting for linear and non-linear effects.

At 260, an output is provided. For example, the example identification system 140 can provide the example output 150.

At 270 another nucleic acid sample is isolated from another sample from the subject. For example, the example nucleic acid isolation module 110 can be configured to isolate and/or substantially purify nucleic acid compositions from another sample to produce another example nucleic acid sample. Since the collection of genotype data 122 from a subject does not change over time, the newly produced nucleic acid sample can be used to obtain

methylation data 132, which is used along with the existing collection of genotype data 122 to provide an updated output (e.g., to perform a checkup on the subject 101 at a later point in time). In some implementations, this abbreviated process can be performed on a periodic or semi-periodic basis to provide ongoing monitoring of one or more medical conditions identified for the subject 101.

FIG. 3 is a block diagram of example computing devices 300, 350 that may be used to implement the systems and methods described in this document, either as a client or as a server or plurality of servers. Computing device 300 is intended to represent various forms of digital computers, such as laptops, desktops, workstations, personal digital assistants, servers, blade servers, mainframes, and other appropriate computers. Computing device 300 can also represent all or parts of various forms of computerized devices, such as embedded digital controllers, media bridges, modems, network routers, network access points, network repeaters, and network interface devices including mesh network communication interfaces. Computing device 350 is intended to represent various forms of mobile devices, such as personal digital assistants, cellular telephones, smartphones, and other similar computing devices. The components shown here, their connections and relationships, and their functions, are meant to be exemplary only, and are not meant to limit implementations of the compositions and methods described herein.

Computing device 300 includes a processor 302, a memory 304, a storage device 306, a high-speed interface 308 connecting to memory 304 and high-speed expansion ports 310, and a low-speed interface 312 connecting to a low-speed bus 314 and storage device 306. Each of the components 302, 304, 306, 308, 310, and 312, are interconnected using various busses, and may be mounted on a common motherboard or in other manners as appropriate. The processor 302 can process instructions for execution within the computing device 300, including instructions stored in the memory 304 or on the storage device 306 to display graphical information for a GUI on an external input/output device, such as display 316 coupled to high-speed interface 308. In other implementations, multiple processors and/or multiple buses may be used, as appropriate, along with multiple memories and types of memory. Also, multiple computing devices 300 may be connected, with each device providing portions of the necessary operations (e.g., as a server bank, a group of blade servers, or a multi-processor system).

The memory 304 stores information within the computing device 300. In one implementation, the memory 304 is a computer-readable medium. In one implementation, the memory 304 is a volatile memory unit or units. In another implementation, the memory 304 is a non-volatile memory unit or units.

5 The storage device 306 can provide mass storage for the computing device 300. In one implementation, the storage device 306 is a computer-readable medium. In various implementations, the storage device 306 may be a floppy disk device, a hard disk device, an optical disk device, or a tape device, a flash memory or other similar solid state memory device, or an array of devices, including devices in a storage area network or other
10 configurations. In one implementation, a computer program product is tangibly embodied in an information carrier. The computer program product contains instructions that, when executed, perform one or more methods, such as those described above. The information carrier is a computer- or machine-readable medium, such as the memory 304, the storage device 306, or memory on processor 302.

15 The high-speed controller 308 manages bandwidth-intensive operations for the computing device 300, while the low-speed controller 312 manages lower bandwidth-intensive operations. Such allocation of duties is exemplary only. In one implementation, the high-speed controller 308 is coupled to memory 304, display 316 (e.g., through a graphics processor or accelerator), and to high-speed expansion ports 310, which may accept various
20 expansion cards (not shown). In the implementation, low-speed controller 312 is coupled to storage device 306 and low-speed expansion port 317 through the low-speed bus 314. The low-speed expansion port, which may include various communication ports (e.g., Universal Serial Bus (USB), BLUETOOTH, BLUETOOTH Low Energy (BLE), Ethernet, wireless Ethernet (WiFi), High-Definition Multimedia Interface (HDMI), ZIGBEE, visible or infrared
25 transceivers, Infrared Data Association (IrDA), fiber optic, laser, sonic, ultrasonic) may be coupled to one or more input/output devices, such as a keyboard, a pointing device, a scanner, a networking device such as a gateway, modem, switch, or router, e.g., through a network adapter 313.

30 Peripheral devices can communicate with the high-speed controller 308 through one or more peripheral interfaces of the low-speed controller 312, including but not limited to a USB stack, an Ethernet stack, a WiFi radio, a BLUETOOTH Low Energy (BLE) radio, a

ZIGBEE radio, an HDMI stack, and a BLUETOOTH radio, as is appropriate for the configuration of a sensor. For example, a sensor that outputs a reading over a USB cable can communicate through a USB stack.

5 The network adapter 313 can communicate with a network 315. Computer networks typically have one or more gateways, modems, routers, media interfaces, media bridges, repeaters, switches, hubs, Domain Name Servers (DNS), and Dynamic Host Configuration Protocol (DHCP) servers that allow communication between devices on the network and devices on other networks (e.g., the Internet). One such gateway can be a network gateway that routes network communication traffic among devices within the network and devices
10 outside of the network. One common type of network communication traffic that is routed through a network gateway is a Domain Name Server (DNS) request, which is a request to the DNS to resolve a uniform resource locator (URL) or uniform resource indicated (URI) to an associated Internet Protocol (IP) address.

The network 315 can include one or more networks. The network(s) may provide for
15 communications under various modes or protocols, such as Global System for Mobile communication (GSM) voice calls, Short Message Service (SMS), Enhanced Messaging Service (EMS), or Multimedia Messaging Service (MMS) messaging, Code Division Multiple Access (CDMA), Time Division Multiple Access (TDMA), Personal Digital Cellular (PDC), Wideband Code Division Multiple Access (WCDMA), CDMA2000,
20 General Packet Radio System (GPRS), or one or more television or cable networks, among others. For example, the communication may occur through a radio-frequency transceiver. In addition, short-range communication may occur, such as using a BLUETOOTH, BLE, ZIGBEE, WiFi, IrDA, or other such transceiver.

In some embodiments, the network 315 can have a hub-and-spoke network
25 configuration. A hub-and-spoke network configuration can allow for an extensible network that can accommodate components being added, removed, failing, and replaced. This can allow, for example, more, fewer, or different devices on the network 315. For example, if a device fails or is deprecated by a newer version of the device, the network 315 can be configured such that network adapter 313 can be updated about the replacement device.

30 In some embodiments, the network 315 can have a mesh network configuration (e.g., ZIGBEE). Mesh configurations may be contrasted with conventional star/tree network

configurations in which the networked devices are directly linked to only a small subset of other network devices (e.g., bridges/switches), and the links between these devices are hierarchical. A mesh network configuration can allow infrastructure nodes (e.g., bridges, switches, and other infrastructure devices) to connect directly and non-hierarchically to other nodes. The connections can dynamically self-organize and self-configure to route data. By not relying on a central coordinator, multiple nodes can participate in the relay of information. In the event of a failure of one or more of the nodes or the communication links between them, the mesh network can self-configure to dynamically redistribute workloads and provide fault-tolerance and network robustness.

The computing device 300 may be implemented in a number of different forms, as shown in the figure. For example, it may be implemented as a standard server 320, or multiple times in a group of such servers. It may also be implemented as part of a rack server system 324. It may also be implemented as part of network device such a modem, gateway, router, access point, repeater, mesh node, switch, hub, or security device (e.g., camera server). In addition, it may be implemented in a personal computer such as a laptop computer 322. Alternatively, components from computing device 300 may be combined with other components in a mobile device (not shown), such as device 350. In some embodiments, the device 350 can be a mobile telephone (e.g., a smartphone), a handheld computer, a tablet computer, a network appliance, a camera, an enhanced general packet radio service (EGPRS) mobile phone, a media player, a navigation device, an email device, a game console, an interactive or smart television, a media streaming device, or a combination of any two or more of these data processing devices or other data processing devices. In some implementations, the device 350 can be included as part of a motor vehicle (e.g., an automobile, an emergency vehicle (e.g., fire truck, ambulance), a bus). Each of such devices may contain one or more of computing device 300, 350, and an entire system may be made up of multiple computing devices 300, 350 communicating with each other through a low-speed bus or a wired or wireless network.

Computing device 350 includes a processor 352, memory 364, an input/output device such as a display 354, a communication interface 366, and a transceiver 368, among other components. The device 350 may also be provided with a storage device, such as a microdrive or other device, to provide additional storage. Each of the components 350, 352,

364, 354, 366, and 368, are interconnected using various buses, and several of the components may be mounted on a common motherboard or in other manners as appropriate.

The processor 352 can process instructions for execution within the computing device 350, including instructions stored in the memory 364. The processor may also include
5 separate analog and digital processors. The processor may provide, for example, for coordination of the other components of the device 350, such as control of user interfaces, applications run by device 350, and wireless communication by device 350.

Processor 352 may communicate with a user through control interface 358 and display interface 356 coupled to a display 354. The display 354 may be, for example, a TFT
10 LCD display or an OLED display, or other appropriate display technology. The display interface 356 may comprise appropriate circuitry for driving the display 354 to present graphical and other information to a user. The control interface 358 may receive commands from a user and convert them for submission to the processor 352. In addition, an external interface 362 may be provide in communication with processor 352, so as to enable near area
15 communication of device 350 with other devices. External interface 362 may provide, for example, for wired communication (e.g., via a docking procedure) or for wireless communication (e.g., via Bluetooth or other such technologies).

The memory 364 stores information within the computing device 350. In one implementation, the memory 364 is a computer-readable medium. In one implementation, the
20 memory 364 is a volatile memory unit or units. In another implementation, the memory 364 is a non-volatile memory unit or units. Expansion memory 374 may also be provided and connected to device 350 through expansion interface 372, which may include, for example, a SIMM card interface. Such expansion memory 374 may provide extra storage space for device 350 or may also store applications or other information for device 350. Specifically,
25 expansion memory 374 may include instructions to carry out or supplement the processes described above and may include secure information also. Thus, for example, expansion memory 374 may be provide as a security module for device 350 and may be programmed with instructions that permit secure use of device 350. In addition, secure applications may be provided via the SIMM cards, along with additional information, such as placing
30 identifying information on the SIMM card in a non-hackable manner.

The memory may include for example, flash memory and/or MRAM memory, as discussed below. In one implementation, a computer program product is tangibly embodied in an information carrier. The computer program product contains instructions that, when executed, perform one or more methods, such as those described above. The information carrier is a computer- or machine-readable medium, such as the memory 364, expansion memory 374, or memory on processor 352.

Device 350 may communicate wirelessly through communication interface 366, which may include digital signal processing circuitry where necessary. Communication interface 366 may provide for communications under various modes or protocols, such as GSM voice calls, Voice Over LTE (VOLTE) calls, SMS, EMS, or MMS messaging, CDMA, TDMA, PDC, WCDMA, CDMA2000, GPRS, WiMAX, LTE, 5G, among others. Such communication may occur, for example, through radio-frequency transceiver 368. In addition, short-range communication may occur, such as using a Bluetooth, WiFi, or other such transceiver (not shown) configured to provide uplink and/or downlink portions of data communication. In addition, GPS receiver module 370 may provide additional wireless data to device 350, which may be used as appropriate by applications running on device 350.

Device 350 may also communicate audibly using audio codec 360, which may receive spoken information from a user and convert it to usable digital information. Audio codec 360 may likewise generate audible sound for a user, such as through a speaker, e.g., in a handset of device 350. Such sound may include sound from voice telephone calls, may include recorded sound (e.g., voice messages, music files, etc.) and may also include sound generated by applications operating on device 350.

The computing device 350 may be implemented in a number of different forms, as shown in the figure. For example, it may be implemented as a cellular telephone 380. It may also be implemented as part of a smartphone 382, personal digital assistant, or other similar mobile device.

Various implementations of the systems and techniques described here can be realized in digital electronic circuitry, integrated circuitry, specially designed ASICs (application specific integrated circuits), computer hardware, firmware, software, and/or combinations thereof. These various implementations can include implementation in one or more computer programs that are executable and/or interpretable on a programmable system

including at least one programmable processor, which may be special or general purpose, coupled to receive data and instructions from, and to transmit data and instructions to, a storage system, at least one input device, and at least one output device.

These computer programs (also known as programs, software, software applications
5 or code) include machine instructions for a programmable processor and can be implemented in a high-level procedural and/or object-oriented programming language, and/or in assembly/machine language. As used herein, the terms “machine-readable medium” “computer-readable medium” refers to any computer program product, apparatus and/or device (e.g., magnetic discs, optical disks, memory, Programmable Logic Devices (PLDs))
10 used to provide machine instructions and/or data to a programmable processor, including a machine-readable medium that receives machine instructions as a machine-readable signal. The term “machine-readable signal” refers to any signal used to provide machine instructions and/or data to a programmable processor.

To provide for interaction with a user, the systems and techniques described here can
15 be implemented on a computer having a display device (e.g., a CRT (cathode ray tube) or LCD (liquid crystal display) monitor) for displaying information to the user and a keyboard and a pointing device (e.g., a mouse or a trackball) by which the user can provide input to the computer. Other kinds of devices can be used to provide for interaction with a user as well; for example, feedback provided to the user can be any form of sensory feedback (e.g., visual
20 feedback, auditory feedback, or tactile feedback); and input from the user can be received in any form, including acoustic, speech, or tactile input.

The systems and techniques described here can be implemented in a computing system that includes a back end component (e.g., as a data server), or that includes a middleware component (e.g., an application server), or that includes a front end component
25 (e.g., a client computer having a graphical user interface or a Web browser through which a user can interact with an implementation of the systems and techniques described here), or any combination of such back end, middleware, or front end components. The components of the system can be interconnected by any form or medium of digital data communication (e.g., a communication network). Examples of communication networks include a local area
30 network (“LAN”), a wide area network (“WAN”), and the Internet.

Some communication networks can be configured to carry power as well as information on the same physical media. This allows a single cable to provide both data connection and electric power to devices. Examples of such shared media include power over network configurations in which power is provided over media that is primarily or previously used for communications. One specific embodiment of power over network is Power Over Ethernet (POE) which pass electric power along with data on twisted pair Ethernet cabling. Examples of such shared media also include network over power configurations in which communication is performed over media that is primarily or previously used for providing power. One specific embodiment of network over power is Power Line Communication (PLC) (also known as power-line carrier, power-line digital subscriber line (PDSL), mains communication, power-line telecommunications, or power-line networking (PLN), Ethernet-Over-Power (EOP)) in which data is carried on a conductor that is also used simultaneously for AC electric power transmission.

The computing system can include clients and servers. A client and server are generally remote from each other and typically interact through a communication network. The relationship of client and server arises by virtue of computer programs running on the respective computers and having a client-server relationship to each other.

The computing system can include routers, gateways, modems, switches, hub, bridges, and repeaters. A router is a networking device that forwards data packets between computer networks and performs traffic directing functions. A network switch is a networking device that connects networked devices together by performing packet switching to receive, process, and forward data to destination devices. A gateway is a network device that allows data to flow from one discrete network to another. Some gateways can be distinct from routers or switches in that they can communicate using more than one protocol and can operate at one or more of the seven layers of the open systems interconnection model (OSI). A media bridge is a network device that converts data between transmission media so that it can be transmitted from computer to computer. A modem is a type of media bridge, typically used to connect a local area network to a wide area network such as a telecommunications network. A network repeater is a network device that receives a signal and retransmits it to extend transmissions and allow the signal can cover longer distances or overcome a communications obstruction.

It will be apparent that the present disclosure provides a skilled artisan the ability to construct a matrix in which the methylation status of one or more CpG dinucleotides and/or one or more genotypes (e.g., SNPs; e.g., at one or more alleles) can be evaluated as described herein, typically using a computer, to identify interactions and allow for prediction of the presence or incidence of CVD. Although such an analysis is complex, no undue experimentation is required as all necessary information is either readily available to the skilled artisan or can be acquired by experimentation as described herein.

Methods of Treating, Managing and/or Monitoring Cardiovascular Diseases

The present disclosure provides methods for determining the likelihood that a subject has CVD, methods for monitoring a subject for CVD (e.g., progression of disease), methods of determining the severity of the CVD (e.g., degree of obstruction), and/or estimating the survival of a subject at risk for or who has CVD. As used herein, CVD includes, without limitation, CHD, stroke, arrhythmia, cardiac arrest, and congestive heart failure. The methods and compositions described herein provide a better ability to assess a subject's risk for or monitor the presence of cardiovascular disease, which is the first step toward more effective prevention. In addition, the methods and compositions described herein provide the ability to estimate the survival of a subject at risk for or who has been determined to have CVD, which can allow for therapies and/or lifestyle changes that may prolong or extend the survival of the subject.

Upon making a positive prognosis of a cardiac outcome (e.g., a prognosis of cardiovascular death, myocardial infarct (MI), stroke, all cause death, or a composite thereof), a medical practitioner can advantageously use the prognostic information thereby obtained to identify the need for, or to customize or optimize, an intervention in the subject, such as, for example, stress testing with ECG response or myocardial perfusion imaging, coronary computed tomography angiogram, diagnostic cardiac catheterization, percutaneous coronary (e.g., balloon angioplasty with or without stent placement), coronary artery bypass graft (CABG), enrollment in a clinical trial, and administration or monitoring of effects of agents selected from, but not limited to, of agents selected from nitrates, beta blockers, ACE inhibitors, antiplatelet agents and lipid-lowering agents. In addition, a medical practitioner can advantageously use the prognostic information thereby obtained to make

recommendations for lifestyle changes including, without limitation, diet modification, exercise regimens, smoking and/or drinking cessation, and combinations thereof. Using the information provided by the methods described herein, a medical practitioner can manage the interventions and monitor an individual to observe, e.g., an improvement.

Those identified as being at higher risk (e.g., PPV of 69% for CVD) or as having the disease can be followed up promptly for further testing or more aggressive interventions. Conversely, those at lower risk can be re-tested periodically and monitored to ensure continued prevention due to the dynamic nature of DNA methylation.

Interventions for cardiovascular disease can depend on the type of cardiovascular disease and the symptoms the individual is experiencing. Interventions for cardiovascular disease can be preventative, therapeutic or palliative. Treatments for cardiovascular diseases can include, for example, lifestyle changes (e.g., diet (e.g., low fat diet), weight loss, exercise, reduction or cessation in smoking and/or drinking), therapeutics (e.g., beta blockers, statins, calcium channel blocker, ACE inhibitors, vasodilator, alteplase, small molecule modulators, pre-/pro-/syn-/post-biotics), medical interventions (e.g., angioplasty, bypass surgery, implantable device, endarterectomy), gene therapy, gene editing, base editing, epigenetic therapy, epigenetic silencing, and/or epigenetic editing.

In accordance with the present disclosure, there may be employed conventional molecular biology, microbiology, biochemical, and recombinant DNA techniques within the skill of the art. Such techniques are explained fully in the literature. The invention will be further described in the following examples, which do not limit the scope of the methods and compositions of matter described in the claims.

25 EXAMPLES

Example 1—Materials and Methods

This study features data and/or biomaterial from three sources. The first set of anonymized genome-wide genetic, genome-wide DNA methylation and clinical data are from the Framingham Heart Study (FHS) Offspring cohort, the second set of anonymized clinical data and DNA are from the Intermountain Healthcare (IM) biorepository, and the third set is from an Iowa cohort as described in more detail below. The procedures and

protocols used for the analysis of the FHS data and the Iowa cohort were approved by the University of Iowa Institutional Review Board (IRB# 201503802 and IRB# 201910834), and the procedures and protocols used for the analyses of the IM materials were approved by the Intermountain Healthcare Institutional Review Board (IRB# 1024811).

5

Example 2—Framingham Heart Study (FHS) Offspring Cohort

The details on the collection and preparation of clinical and biological data of the FHS cohort have been described previously (dbGAP study accession: phs000007). In brief, the demographics, risk factors and clinical information were derived from the Offspring cohort, including coronary heart disease (CHD) status. CHD was considered present if an individual was diagnosed with CHD. Conversely, CHD was considered absent if an individual was not diagnosed with CHD. Sources of clinical data in determining CHD events included subject report, review of medical records, and death certificates. The designations and dates of CHD onset used in this study are as determined by a panel of three investigators on the Framingham Endpoint Review Committee, but could similarly be applied to other CVDs.

Genome-wide DNA methylation data profiled using the Illumina Infinium HumanMethylation450 BeadChip array (San Diego, CA, USA) was available from 2,567 subjects who were phlebotomized. Standard sample and probe level quality control were performed as described in previous studies, which resulted in retaining 2,560 samples and DNA methylation data from 403,192 loci (see, e.g., Dogan et al., 2018, Genes, 9:641; Pidsley et al., 2013, BMC Genomics, 14:1-10; Triche, 2014, FDb.InfiniumMethylation.hg19: Annotation package for Illumina Infinium DNA methylation probes. Vol. R package version 2.2.0; Davis et al., 2018, Handle Illumina methylation data., Vol. R package version 2.22.0; and Dogan et al., 2018, PLoS One, 13:e0190549). Genome-wide genotype data obtained using the Affymetrix GeneChip HumanMapping 500K array (Santa Clara, CA, USA) was available for 2,406 of the remaining samples. After standard sample and probe level quality control procedures were performed in PLINK on the array data as described previously, the total number of samples and SNPs remaining were 2,295 and 472,822, respectively (Dogan et al., 2018, Genes, 9:641; Dogan et al., 2018, PLoS One, 13:e0190549; and Purcell et al., 2007, Am. J. Hum. Genet., 81:559-75). Based on the number of those diagnosed with CHD

30

(cases) and those that were not diagnosed with or did not have a CHD event within four years of examination (controls), the total number of subjects was 2,111. The demographics and conventional risk factors of these individuals are summarized in Table 1.

5 Table 1. Summary of demographics and conventional CHD risk factors for the individuals in the Framingham Heart Study Offspring cohort

	FHS Training (n = 1583)						FHS Test (n = 528)					
	No CHD			CHD			No CHD			CHD		
Gender (count)												
Females	812			60			271			20		
Males	588			123			196			41		
Age (years)												
Females	65.8	±	8.9	73.2	±	8.8	66.0	±	8.4	71.8	±	8.8
Males	64.9	±	8.7	70.0	±	7.9	64.8	±	8.5	70.1	±	7.9
Total cholesterol (mg/dl)												
Females	197.6	±	35.1	173.6	±	35.5	200.6	±	36.3	176.4	±	40.0
Males	179.1	±	32.5	151.0	±	32.9	176.8	±	31.9	147.0	±	24.4
HDL cholesterol (mg/dl)												
Females	64.8	±	18.6	60.0	±	18.8	66.1	±	19.2	58.1	±	12.8
Males	50.5	±	13.9	45.6	±	11.4	49.9	±	14.6	45.0	±	11.1
HbA1c (%)												
Females	5.7	±	0.6	6.1	±	0.9	5.6	±	0.4	6.0	±	0.9
Males	5.7	±	0.7	6.0	±	0.9	5.7	±	0.8	6.0	±	1.0
SBP (mmHg)												
Females	127.7	±	17.6	135.8	±	16.7	129.0	±	18.4	132.1	±	19.8
Males	129.6	±	16.7	124.4	±	18.3	128.4	±	16.5	133.1	±	19.4
DBP (mmHg)												
Females	72.7	±	9.9	69.2	±	10.9	73.6	±	10.2	67.8	±	12.0
Males	76.5	±	10.3	68.8	±	10.9	75.9	±	10.4	70.3	±	13.4

10 HDL: high-density lipoprotein, HbA1c: Hemoglobin A1c, SBP: systolic blood pressure, DBP: diastolic blood pressure.

Example 3—Intermountain Healthcare Cohort

15 The first independent validation cohort consisted of 252 subjects from the Intermountain Healthcare (IM) Heart Institute INSPIRE registry who underwent coronary angiography. A CHD case subject was defined as an adult >18 years old whom did not have a history of CHD or myocardial infarction (MI) prior to the index coronary angiogram but had a clinical diagnosis of CHD (>70% stenosis) on angiography. A control subject was

defined as an adult >18 years old whom did not have a history of CHD or myocardial infarction (MI) prior to the index coronary angiogram, had no clinical diagnosis of CHD (<50% stenosis) at the index coronary angiography and no clinical diagnosis of CHD (>70% stenosis) on angiography, MI, revascularization, or death due to CHD within four years of index coronary angiography.

The demographics of these individuals are summarized in Table 2.

Table 2. Summary of demographics for the Intermountain Healthcare validation sets

	CHD cases (n=127)	CHD controls (n=125)
Sex: male (count)	59	58
Sex: female (count)	68	67
Age (years)	66.0 ± 14.2	63.3 ± 15.3

Genome-wide DNA methylation and genetic assessments for each of these 253 subjects were conducted by the University of Minnesota Genome Center using the Illumina Infinium MethylationEpic Beadchip array and the Illumina Infinium Multi-Ethnic Global BeadChip array (San Diego, CA, USA), respectively. These data were then subjected to the same quality control procedure described above for the FHS samples.

Example 4—Iowa Cohort

The second independent validation cohort consisted of 167 subjects. The demographics are shown in Table 3. The presence or absence of a clinical diagnosis of CHD was through medical records.

Table 3. Summary of demographics for the Iowa validation set

	CHD cases (n=978)	CHD controls (n=47)
Sex: male (count)	57	16
Sex: female (count)	21	31
Age (years)	64.7 ± 13.0	48.6 ± 11.0

Example 5—Integrated Genetic-Epigenetic Coronary Heart Disease Risk Prediction Model

Because one of the aims of this study is to translate array-based methylation loci to clinically implementable digital PCR (dPCR) assays, which has fixed constraints on precision, prior to performing data mining exclusively using data from the FHS training set, the methylation variables were reduced to include loci based on delta beta ($\Delta\beta$) (absolute difference between case and controls). All methylation loci beta values were converted into M-values and subsequently scaled to have zero mean and unit variance.

All data mining, feature selection, model development and model tuning were performed exclusively on the FHS training set. Our data mining approach has been outlined in previous publications (Dogan et al., 2018, Genes, 9:641; Dogan et al., 2018, PLoS One, 13:e0190549). All analyses were performed in Python. Briefly, an undersampling-based approach was implemented to account for the high class imbalance and coupled to an ensemble of machine learning algorithms that incorporated cross-validation to uncover non-linear methylation-SNP interactions and highly predictive biosignatures in the FHS training set (Han et al., 2011, *Data Mining: Concepts and Techniques*, Elsevier). As a result, a marker set was selected consisting of six DNA methylation loci and ten SNPs that had the best combined performance with respect to area under the receiver operating characteristic curve (AUC), sensitivity and specificity. The ensemble model consisting of these 16 biomarkers underwent hyperparameter tuning and was finalized for testing.

Example 6—Survival Analysis and Prognostic Scores

Using data from the FHS, a Kaplan-Meier survival curve and Cox Proportional Hazards can be fitted to display CHD as a function of risk group (high vs. low) as predicted by the integrated genetic-epigenetic model. The y-axis represents the probability of not having CHD. The 95% confidence interval (CI) for each of the distribution was calculated and the distributions of the high and low risk groups were compared using the log-rank test.

Example 7—Results

The clinical and demographic characteristics of the FHS, IM and Iowa cohorts are outlined in Tables 1, 2, and 3, respectively. All of the subjects from the FHS cohort were of

European ancestry, but non-European ancestry was represented in the IM and Iowa cohorts. The most notable difference was with respect to gendering. Compared to the FHS and IM cohorts, the CHD controls (those not diagnosed with CHD) were younger in the Iowa cohort.

5 Example 8—Integrated Genetic-Epigenetic Coronary Heart Disease Risk Prediction Model

Using integrated genome-wide SNP and methylation data from the training sets, a CHD prediction model was built to identify those that have CHD. All subjects had genetic (SNPs) and epigenetic (DNA methylation) molecular data. All data mining, variable selection, and model development work were performed on the FHS training set. The data
10 from the FHS test set, and IM and Iowa independent external validation sets were used to validate the performance of the final model developed using the FHS training set. Using the data from the FHS training set, machine learning (a subset of artificial intelligence) procedures were used to develop a model for the detection of CHD. The final model was built using data from ten SNPs and six DNA methylation loci, for a total of 16 biomarkers.
15 The performance of that model was then tuned and upon finalization, was independently examined in the FHS test and IM and Iowa independent external validation sets to better understand the generalizability of this biomarker panel.

This final ensemble model consisted of a total of 16 biomarkers, six of which were DNA methylation biomarkers and the remaining ten were SNPs. The six methylation loci are
20 cg04988978 (5' promoter region of MPO), cg21161138 (gene body of AHRH), cg12655112 (gene body of EHD4), cg03725309 (body of SARS1), cg12586707 (3' intergenic region of CXCL1), and cg17901584 (gene body of DHCR24), while the ten SNPs are rs2869675 (gene body of PREX1), rs4376434 (intergenic region near LINC00972), rs12129789 (gene body of KCND3), rs7585056 (intergenic region near TMEM18), rs710987 (gene body of
25 LINC010019), rs4639796 (gene body of ZBTB41), rs1333048 (3' intergenic region of CDKN2B), rs12714414 (intergenic region near TMEM18), rs942317 (gene body of KTN1-AS1), and rs1441433 (gene body of PPP3CA).

The overall and sex-specific performances of this model for the detection of CHD across all three cohorts are shown in Table 4. As expected, PrecisionCHD had the best
30 performance in the FHS training data set which was used to develop the model. More importantly, PrecisionCHD demonstrated robust generalizability. It had 75% or better

sensitivity across all cohorts, with the highest validation sensitivity and specificity being 88% and 77% in the external Iowa validation cohort. Across the three sets (i.e. FHS test, IM and Iowa) not used in the training of the model, overall, the model performed with an average AUC, sensitivity and specificity of 81%, 80% and 75%, respectively. A 80% sensitivity (true positive rate) indicates that, of 100 individuals with CHD, 80 are identified correctly by PrecisionCHD. Similarly, a 75% specificity (true negative rate) indicates that, of 100 without CHD, 75 are identified correctly. Similarly, the average sensitivity and specificity for men were 81% and 73%, respectively. For women, the average sensitivity and specificity were 76% and 75%, respectively. Overall, the model performed similarly for both men and women and across cohorts, indicating minimal to no gender bias and robust generalizability.

Table 4. Performance of PrecisionCHD™ in the Framingham Heart Study, Intermountain Healthcare and Iowa cohorts.

Dataset	AUC	Sensitivity	Specificity
FHS Test			
Overall	0.82	0.77	0.75
Male	0.81	0.75	0.73
Female	0.83	0.78	0.77
IM Validation			
Overall	0.73	0.75	0.71
Male	0.75	0.76	0.70
Female	0.72	0.74	0.70
Iowa Validation			
Overall	0.86	0.88	0.77
Male	0.83	0.92	0.75
Female	0.82	0.76	0.77

AUC: area under the receiver operating characteristic curve
FHS: Framingham Heart Study cohort
IM: Intermountain Healthcare cohort
Iowa: Iowa cohort
Sensitivity: true positive rate; Specificity: true negative rate

Example 9—Additional Data

Appendix A shows a list of CpGs whose methylation is associated with CVD. Appendix B shows a list of genes whose methylation is associated with CVD. Appendix C shows a list of SNPs associated with CVD. The numerical values provided in Appendix A, B,

and C are the mean of 10-fold cross validation scores, AUC ROC (Area Under The Receiver Operating Characteristic Curve), sensitivity and specificity, which were computed by logistic regression. Sensitivity is the true positive rate and specificity is the true negative rate.

5 Example 10—PrecisionCHD

 PrecisionCHD is a quantitative test to aid in the early detection of coronary heart disease (CHD). See Table 5. This non-invasive test evaluates ten single nucleotide polymorphisms (SNPs) and six DNA methylation markers in genomic DNA isolated from human peripheral whole blood. Coronary heart disease results from heritable (genetic) and
10 acquired, potentially modifiable lifestyle and environmental (epigenetic) factors. The PrecisionCHD early detection test measures certain complex genetic and epigenetic relationships associated with CHD and uses a machine learning model to predict CHD status.

 The PrecisionCHD test is initially intended for adults between the ages of 35-80 presenting to be evaluated for coronary heart disease. The results of this test are intended to
15 be interpreted by a healthcare provider in conjunction with a comprehensive medical evaluation. This test is not indicated for stand-alone coronary heart disease diagnostic purposes and is not intended to replace a healthcare professional's diagnosis and treatment of coronary heart disease.

20 Table 5. Summary of PrecisionCHD

Clinical purpose of the test	Detection of coronary heart disease
Target drug(s), if applicable	None
Target condition	Coronary heart disease
The population for which the test is intended	Adults between the ages of 35-80 presenting to be evaluated for coronary heart disease
Test class	Cardiovascular disease
Test type	qPCR and digital PCR
Analyte being measured	DNA
Type of specimen(s) acceptable for testing	Whole blood
Output of test results for clinical decision making	Quantitative (CHD signal detected/not detected) Actionable Clinical Intelligence platform mapping biomarkers to key drivers of CHD
Board-certified specialist needed for test result interpretation?	No

As described herein, PrecisionCHD evaluates a total of 16 biomarkers, including ten SNP genotypes and six DNA methylation biomarkers. The PrecisionCHD test uses standard Taqman assays to profile genotypes and proprietary methylation sensitive digital PCR assays to profile DNA methylation markers. The biomarkers captured by PrecisionCHD map to several complex pathways associated with the biology and pathogenesis of CHD such as serine metabolism, cholesterol biosynthesis, smoking and inflammation. Cardio Diagnostics' Actionable Clinical Intelligence™ (ACI™) platform (see, for example, US Application No. 63/488,463, incorporated herein by reference) maps biomarker information to modifiable risk factors for CHD to help guide personalized interventions.

Example 11—PrecisionCHD Workflow

Undergoing testing with PrecisionCHD is simple, fast, and convenient. The steps include:

1. Eligibility criteria:
 - a. 35-80 years old
 - b. Presenting to be evaluated for coronary heart disease
 - i. Exclusion: patients that have undergone a bone marrow transplant are not eligible
2. Sample collection:
 - a. Option 1: at-home lancet-based sample collection kit mailed directly to the patient upon receiving test order from a clinician.
 - b. Option 2: blood draw in provider setting.
 - c. Option 3: blood draw in a non-provider setting such as a mobile clinic, community center, phlebotomy center.
3. Sample processed at high-complexity CLIA lab to profile genotype and methylation biomarkers.
4. Analytics of biomarkers are performed and a clinical report is generated and shared with the ordering clinician.
5. Clinicians will also receive a login to the Actionable Clinical Intelligence platform (see, for example, US Application No. 63/488,463, incorporated herein

by reference) with supplementary mapping of molecular information to modifiable risk factors associated with the patient's status.

6. Clinician's discuss results with patient and prevention/management plan is outlined. FIG. 4 outlines several potential action items that can be implemented based on a positive CHD signal using the PrecisionCHD test.

Example 12—Tailoring Interventions

The methods described herein can assist in selecting and/or customizing an appropriate intervention. The effectiveness of an intervention such as lifestyle or pharmaceutical can be evaluated by administering the test before and after said intervention to quantify changes, if any, in the status or risk to inform other future interventions. This process can be repeated to identify optimal interventions for the individual patient. Therefore, the methods described herein can aid in managing the interventions for an individual patient and then subsequently monitoring the patient to evaluate the efficacy of such interventions.

Example 13—PrecisionCHD-Epi for Assessing Mortality Risk in Those with Coronary Heart Disease

PrecisionCHD™ is a powerful integrated genetic-epigenetic diagnostic tool for detecting the presence of coronary heart disease (CHD) in clinical populations or for life insurance post-issue health initiatives. Its genetics-free version, PrecisionCHD-Epi, is a sensitive screening tool for use by life insurance carriers to screen individuals for the presence of CHD during the underwriting phase. However, if CHD presence has already been established for an individual, we performed experiments to determine if PrecisionCHD-Epi can help further stratify the severity of CHD as it relates to mortality risk.

This is an important question for life insurers, as stratifying mortal risk in those already diagnosed with CHD is one of the highest risk endeavors for medical underwriters. Seeking to put these assessments on a firm medical basis, in an influential article published twenty years ago, Dr. Anthony Milano reviewed the extant literature and recommended using CHD class and left ventricular function as the linchpins of the mortal risk assessment (Milano, 2000, J. Insur. Med. New York, 32(3):167-85). Since that time, other markers of CHD mortality, such as N-terminal Pro Brain Natriuretic Protein (NT-ProBNP), have been

added to the underwriting armamentarium to aid prediction (Bibbins-Domingo et al., 2007, JAMA, 297(2):169-76). Still, prediction of mortality in those with CHD is still not optimal. In 2009, Sijbrands and colleagues used the underwriting procedures of Nationale-
 5 3,963 subjects with current cardiovascular disease (CVD) (Sijbrands et al., 2009, PloS One, 4(5):e5457). Despite the complete availability of medical information, and the clear relationship between CVD and mortal risk, Sijbrands concluded that “decedents could not have been identified individually by the medical evaluation employed.”

PrecisionCHD-Epi may be able change this dynamic and help underwriters stratify
 10 mortal risk from CHD. PrecisionCHD-Epi uses the same six methylation sensitive digital PCR assays that power PrecisionCHD™ to sensitively screen for the molecular signature of CHD. Inherent in this approach is the understanding that not all forms of CHD are alike, and that both the treatment and prognoses of patients determined to have CHD will differ.

In order to demonstrate that point and show the power of this technology to predict
 15 CHD severity, we analyzed the clinical and epigenetic data from 244 subjects in the Framingham Heart Study (FHS) who were diagnosed by the FHS Endpoint Committee as having CHD at Wave 8 of the FHS Offspring Cohort Study (Cupples et al., 1988, “The Framingham Heart Study, Section 35. An Epidemiological Investigation of Cardiovascular Disease Survival following Cardiovascular Events: 30 Year Follow-up,” Lung and Blood
 20 Institute). Table 6 delineates the characteristics of the subjects. Underscoring the lack of recognition of CHD in women by current methods, two-thirds of those diagnosed with CHD were male. The rate of self-reported smoking was higher in the males, but low overall. The values for other parameters such as blood pressure and lipid values, were clinically unremarkable.

Table 6. Clinical and Demographic Characteristics

	Male	Female
N	164	80
N Died in Follow Up	48 (29%)	23 (29%)
Age	70 ± 8	73 ± 9
Smokers	18 (11%)	3 (4%)
HbA1c(%)	6.0 ± 0.9	6.1 ± 0.9
Total Chol (mg/dl)	150 ± 32	174 ± 36

HDL Chol (mg/dl)	45 ± 11	59 ± 17
Triglycerides (mg/dl)	126 ± 88	137 ± 69
Systolic BP (mm Hg)	127 ± 19	135 ± 17
Diastolic BP (mm Hg)	69 ± 12	69 ± 11

Table 7 shows the non-digitally transformed methylation values for each of the six CpG sites targeted by the MSdPCR battery. Males had significantly lower methylation values at cg12655112 and cg12586707, with demethylation at each site being associated with CHD status.

Table 7. MSdPCR DNA Methylation Values of PrecisionCHD-Epi Sites by Gender

	Male	Female
cg04988978	17.2 ± 3.8%	18.1 ± 3.9%
cg21161138	84.8 ± 3.2%	85.5 ± 2.9%
cg12655112*	67.3 ± 3.3%	68.3 ± 4.1%
cg03725309	6.8 ± 1.7%	6.5 ± 1.6%
cg12586707**	12.6 ± 3.2%	13.8 ± 3.7%
cg17901584	43.5 ± 6.6%	45.1 ± 7.7%

* p< 0.05, ** p<0.01

We then examined the relationship of the methylation markers to survival. A total of 71 FHS subjects (48 males and 23 females) with CHD were observed as dying in the follow up period. Using their survival times and standard proportional hazards modeling, we analyzed the relationship between methylation values at the six sites to survival using stepwise regression. A simple proportional hazards model constructed using all six MSdPCR markers predicted survival (p<0.006). Neither the addition of age, sex or any of the traditional predictors (lipids, HbA1c and BP) given in Table 1 improved prediction. Stepwise removal of non-significant markers resulted in a model of three markers, cg04988978, cg21161138 and cg12655112, that predicted survival Chi square of p-value of p<0.002 with logWorth values (-log of the p-value) of 2.00, 1.74 and 3.07, respectively.

Visualizing the effect of markers inside an interactive multivariate model of survival can be challenging. An alternative, yet less powerful method of understanding the relationship of methylation to survival can be had by simply plotting the relationship of whether or not the subject died during follow-up to methylation. FIGs. 5A, 5B and 5C illustrate this relationship for each of the three markers to mortal status. Overall, those in the

lower two deciles of cg04988978 methylation were four times (28 of 50 vs 7 of 50) more likely to die during the follow-up period than those in the two highest deciles. Similar, yet less powerful, effects were observed for methylation at cg12655112 and cg21161138.

Conversely, among those who died, there was no significant relationship between cg04988978 and days survived using bivariate analyses as shown in FIG. 6A. However, for both cg12655112 and cg21161138, there were strong significant relationships between days survived and methylation status using bivariate analysis as shown in FIGs. 6B and 6C. Taken together, these suggest a strong, yet complex relationship between methylation markers in the PrecisionCHD-Epi test and survival.

In summary, using either PrecisionCHD or PrecisionCHD-Epi, we can now provide survival estimates for those at risk of developing CHD or those determined to have CHD.

It is to be understood that, while the methods and compositions of matter have been described herein in conjunction with a number of different aspects, the foregoing description of the various aspects is intended to illustrate and not limit the scope of the methods and compositions of matter. Other aspects, advantages, and modifications are within the scope of the following claims.

Disclosed are methods and compositions that can be used for, can be used in conjunction with, can be used in preparation for, or are products of the disclosed methods and compositions. These and other materials are disclosed herein, and it is understood that combinations, subsets, interactions, groups, etc. of these methods and compositions are disclosed. That is, while specific reference to each various individual and collective combinations and permutations of these compositions and methods may not be explicitly disclosed, each is specifically contemplated and described herein. For example, if a particular composition of matter or a particular method is disclosed and discussed and a number of compositions or methods are discussed, each and every combination and permutation of the compositions and the methods are specifically contemplated unless specifically indicated to the contrary. Likewise, any subset or combination of these is also specifically contemplated and disclosed.

Appendix A

ILMNID	mean_cv_auc	GENE	CHR	MAPINFO	ILMNID	mean_cv_auc	GENE	CHR	MAPINFO
cg00001793	0.66426667	ETV6	12	11905390	cg13784312	0.658735	RAPGEF1	9	134609065
cg000008819	0.662035333	C4orf44	4	3257791	cg13788625	0.662511667		9	140340778
cg00010046	0.574812	MGC23284	16	88729480	cg13794530	0.709667	VIPR2	7	158937494
cg00025211	0.659517667	NUDT17	1	145588627	cg13796381	0.679134333	PTGDS	9	139872395
cg00028918	0.546906333		5	159200479	cg13800586	0.515129333	DTYMK	2	242626324
cg00031063	0.701268	MFAP2	1	17303752	cg13814708	0.736352333	ZCCHC24	10	81144099
cg00033551	0.532168333	MGRN1	16	4738568	cg13827209	0.524480333	TGFBR1	9	101912842
cg00061769	0.744212667	YY1AP1	1	155658843	cg13841901	0.738642	MUTYH	1	45805975
cg00065905	0.654327				cg13842237	0.692538333		21	47064079
cg00095214	0.655042333	C1orf122	1	38272200	cg13846507	0.729995	ST6GALNAC4	9	130679184
cg00097968	0.688241667	RPTN	1	152126975	cg13847987	0.662283	ABL1	9	133759967
cg00101940	0.658726667		16	85479301	cg13854219	0.660517		1	101757037
cg00115639	0.663448		17	54701692	cg13865502	0.595623	NRARP	9	140196259
cg00134843	0.735808	PGS1	17	76374495	cg13876222	0.574199667	NOTCH1	9	139399348
cg00142624	0.572734667	FBXO21	12	117627897	cg13879047	0.655835333			
cg00144180	0.6852				cg13880779	0.667552333	NKX2-4	20	21376337
cg00153306	0.680946333	TMIE	3	46741492	cg13905081	0.659286667		9	140199209
cg00154557	0.684007333		6	116763525	cg13919607	0.577879333	ZNF174	16	3451181
cg00156744	0.716290667	STK10	5	171615663	cg13924514	0.658971667		8	142070760
cg00159908	0.677435	LRRC27	10	134184627	cg13961784	0.546672667		1	48175246
cg00166770	0.725227333	C20orf43	20	55043239	cg14004847	0.663278667	MAD1L1	7	1930337
cg00170056	0.712588333	ZNF527	19	37861986	cg14036868	0.751614	ATL2	2	38604442
cg00179423	0.722900667	VPS28	8	145653749	cg14037413	0.739967	ZNF143	11	9482594
cg00208412	0.727830333	GOLGA7B	10	99609481	cg14041100	0.699456667	CHD7	8	61591939
cg00208734	0.654258	SOHLH2	13	36788843	cg14051842	0.684098	LOC100270746	6	26988187
cg00214171	0.657861	RECQL5	17	73642991	cg14051886	0.657456	C20orf114	20	31870206
cg00236005	0.653515667	ZNF365	10	64279749	cg14089007	0.679849	HIP1	7	75164724
cg00237876	0.658566333	CD300LG	17	41924043	cg14102272	0.739653	CLK1	2	201729414
cg00252529	0.67255	ARHGAP10	4	148730174	cg14105458	0.549625333	ABCA2	9	139917436
cg00288562	0.727097667	TRA2A	7	23571645	cg14123992	0.655153333	APOE	19	45407868
cg00316485	0.657380667	ZNF423	16	49789549	cg14139679	0.680633667	SECISBP2	9	91937674
cg00359325	0.627344	USP4	3	49377419	cg14150468	0.564024	SGK3	8	67625464
cg00380464	0.65714	LTV1	6	144164650	cg14153722	0.740458667	MAPKAPK2	1	206858799
cg00393373	0.554108667	ZNF518B	4	10456597	cg14157244	0.723408333	SSB	2	170655474
cg00407659	0.689478333	ANXA6	5	150538414	cg14179148	0.659689		13	40796151
cg00431050	0.669946	ELOVL3	10	103985730	cg14185717	0.518165			
cg00435173	0.656127333	RAB5C	17	40284046	cg14201424	0.535767667	MGC45800	4	183063397
cg00448875	0.704313333	RAB6B	3	133611626	cg14209920	0.655501	NFIX	19	13137642
cg00466121	0.542917	ZNHIT6	1	86174151	cg14217074	0.701543667		3	187528979
cg00473025	0.734321	RBM19	12	114276037	cg14230992	0.690513333		8	16641735
cg00473112	0.536281	RNH1	11	507177	cg14243623	0.734743667	CTSL1	9	90341158
cg00476037	0.525463	TGFA	2	70781307	cg14244402	0.684099667		9	130681102
cg00483459	0.658077	ALS2CL	3	46735782	cg14258623	0.656581667		6	1384591
cg00484466	0.697186	ABLIM2	4	8071202	cg14263139	0.706886667	RIN1	11	66104018
cg00510087	0.697745667				cg14281592	0.578993	ANGPTL2	9	129884313
cg00512454	0.676141333	CPNE5	6	36709214	cg14284211	0.658186	FKBP5	6	35570224
cg00518190	0.661113667	ICMT	1	6294768	cg14329889	0.656254		14	104696094
cg00533827	0.664082		2	86163827	cg14337526	0.662941667	RAP2A	13	98086635
cg00533923	0.748656	RWDD1	6	116892543	cg14341254	0.684235333			
cg00560284	0.674245	SPATS2	12	49783222	cg14397361	0.748909	COBRA1	9	140149997
cg00563932	0.654483667	PTGDS	9	139871049	cg14409810	0.658714	C9orf167	9	140172908
cg00565090	0.664433667	SNORD23	19	48258258	cg14416293	0.663751			
cg00574958	0.550894333	CPT1A	11	68607622	cg14418175	0.664809			
cg00586383	0.675810333		17	75528194	cg14429865	0.674094667	KIAA1688	8	145777710
cg00590670	0.551153333	IGFBP5	2	217559207	cg14455590	0.673186		22	50010191
cg00591471	0.653346	DPRX	19	54135265	cg14476101	0.584862	PHGDH	1	120255992
cg00591980	0.621911	PCMT1	6	150071069	cg14486634	0.671380333		9	98792175
cg00594011	0.726578333	DKK2	4	107956556	cg14488957	0.538727333	ACAP3	1	1244644
cg00602811	0.680070333	ZEB2	2	145278564	cg14507132	0.723898667	TCF12	15	57210852
cg00603498	0.747844	RPAIN	17	5322775	cg14537800	0.717764333	FGFBP2	4	15964950
cg00614832	0.723142333	ZFH3	16	73092394	cg14545834	0.686191333	PTPRD	9	8713437
cg00652657	0.659909	EBF3	10	131761587	cg14580211	0.660395667	C5orf62	5	150161299

Appendix A

cg00657202	0.726528667	MGMT	10	131309164	cg14608275	0.687753	NCEH1	3	172429421
cg00683158	0.595590333	KIAA1543	19	7660969	cg14615927	0.653547	HDAC4	2	240259213
cg00699486	0.518420667		6	166144768	cg14616479	0.628241333	GMPS	3	155588124
cg00703403	0.706618333		17	39810972	cg14624207	0.657335667	LRP5	11	68142198
cg00711896	0.561983333	ZNF48	16	30410051	cg14640386	0.687105333	RABEP1	17	5185735
cg00712390	0.712810333	BAHCC1	17	79373624	cg14649449	0.660223667	MAMDC2	9	72660082
cg00719668	0.684104667		20	34541013	cg14695609	0.690596333	LARS2	3	45430048
cg00720845	0.658589		4	774393	cg14699112	0.680296333		4	141170338
cg00730820	0.670825				cg14753356	0.672344667		6	30720108
cg00734979	0.716120333		10	106004998	cg14795231	0.724088	SIK3	11	116716214
cg00745962	0.671533333	ZFP161	18	5296141	cg14810343	0.739619333	CXXC5	5	139028149
cg00776266	0.661433333	TCP11L2	12	106696255	cg14816748	0.592376	PRPSAP1	17	74349897
cg00777627	0.704213	PDE10A	6	166075755	cg14819942	0.680015667		15	35414228
cg00790835	0.677865	RGS6	14	73029443	cg14826516	0.641419333	SELO	22	50639351
cg00795915	0.747608333	BDP1	5	70751728	cg14830082	0.695851667	ST14	11	130067806
cg00799842	0.518980667	PCIF1	20	44563271	cg14838433	0.707305	TMEM61	1	55446303
cg00803088	0.669753667	RET	10	43600706	cg14860120	0.570010333	MYEF2	15	48470606
cg00805496	0.68858	LOC728640	10	60474759	cg14899046	0.664434667		21	46784677
cg00819141	0.707200667		7	1657422	cg14955151	0.573556333	TRIM47	17	73264635
cg00849642	0.682284	CORO1C	12	109106495	cg14971744	0.601763	ITGAE	17	3627058
cg00850073	0.700103333		5	139016870	cg14978637	0.674457667	EDN2	1	41950817
cg00852675	0.693325333	LOC728743	7	150105086	cg14989202	0.656648667	C1orf200	1	9714843
cg00854594	0.699587		17	36580953	cg15007228	0.658190667	SKI	1	2210335
cg00858400	0.660262333	SLC7A5	16	87904580	cg15014975	0.686724667	RUNX3	1	25257547
cg00863306	0.666396667	NANOS3	19	13991470	cg15048802	0.573757333	C2orf60	2	200820651
cg00869215	0.715026333	CLK2P	7	23626062	cg15062310	0.742776	SLC12A9	7	100450120
cg00871610	0.671978	MIR802	21	37093012	cg15092561	0.662152333	C21orf33	21	45554033
cg00874357	0.621077333	MIRLET7I	12	62997129	cg15167433	0.703845	XRCC1	19	44079985
cg00876267	0.705707		9	139588516	cg15173134	0.666948			
cg00884445	0.673129	VWA3B	2	98755852	cg15174294	0.541736333	ZNF714	19	21265264
cg00923634	0.656906667	MAN1C1	1	25968448	cg15198344	0.710979667	TPRX1	19	48304772
cg00958578	0.652118	SNRNP25	16	103715	cg15256491	0.672172333	FOXK1	7	4729610
cg00975876	0.654858667	PRICKLE4	6	41752769	cg15266508	0.738846333	COL9A3	20	61447742
cg00978910	0.655091	C1orf9	1	172502140	cg15309223	0.732022	TMEM59	1	54519091
cg00988396	0.668053				cg15341833	0.656493667	UXS1	2	106776775
cg00992659	0.617292333	RCC2	1	17766225	cg15356553	0.667011333			
cg00994936	0.530057667	DAZAP1	19	1423902	cg15391574	0.665327333	GLB1L2	11	134201993
cg01016660	0.529145	KIAA1543	19	7660925	cg15407257	0.680767		20	23621136
cg01025283	0.729751	FAIM	3	138327728	cg15408407	0.746436667	RPS15	19	1438438
cg01028796	0.577875333	BICD2	9	95525261	cg15420687	0.654567667	PWWP2A	5	159546321
cg01043250	0.729719667		9	135582139	cg15428620	0.659029667			
cg01054478	0.688937667		5	180100963	cg15461335	0.665855333	SLITRK5	13	88324432
cg01069104	0.714718667				cg15464148	0.654331	LPAR5	12	6745057
cg01079019	0.682116333		16	21566602	cg15502903	0.569777	IL17RA	22	17566297
cg01082498	0.713099667	CPT1A	11	68608225	cg15504677	0.740133667	LHB	19	49521513
cg01102854	0.677970667	MARK2	11	63605717	cg15556709	0.596352333	C7orf44	7	43739176
cg01109287	0.649823	REEP3	10	65281299	cg15593965	0.685288667		8	145844703
cg01121022	0.684698333		14	104338788	cg15601807	0.655434	C10orf25	10	45495677
cg01127300	0.570607667		22	38614796	cg15607292	0.670176333		3	25426810
cg01154537	0.742088	C10orf4	10	95462167	cg15624624	0.659193333	TCF7L2	10	114819357
cg01165142	0.677632	IL4R	16	27367172	cg15644764	0.653904	SYT2	1	202592914
cg01165159	0.654888667				cg15673994	0.680754			
cg01182555	0.72312	HIF1A	14	62162064	cg15697476	0.699223333	TNXB	6	32055474
cg01188574	0.592825	ARID3A	19	925914	cg15700006	0.574699	SKI	1	2160297
cg01191058	0.680669		16	89048433	cg15708175	0.667427333			
cg01234420	0.657652667	LOC150381	22	46453808	cg15736296	0.56914	UBE2L6	11	57335454
cg01240056	0.656449		7	1263927	cg15769920	0.659887	MYL6B	12	56545649
cg01242348	0.580898	CEBPE	14	23586886	cg15774391	0.657754667	STAT5A	17	40439628
cg01245787	0.741043333	DCK	4	71859630	cg15793417	0.658183667	SIPA1L3	19	38573135
cg01257194	0.664420333	ALAD	9	116161247	cg15804598	0.665427667	HEXIM1	17	43224418
cg01266287	0.653850667	RRAD	16	66959833	cg15809217	0.667776333	BAT3	6	31607648
cg01269670	0.739839333	KRBA1	7	149411377	cg15823502	0.670803		6	41650768
cg01275566	0.662429				cg15833447	0.549546667		20	60546782

Appendix A

cg01282725	0.659989667		12	1645963	cg15867698	0.589709	ACTN1	14	69438267
cg01287833	0.676933333	CCHCR1	6	31113026	cg15874885	0.66066			
cg01290904	0.654889333	EVC2	4	5708474	cg15874903	0.708404	SSTR1	14	38678537
cg01305745	0.65453	VKORC1	16	31106788	cg15897774	0.717738667	MTRR	5	7869176
cg01332181	0.733488667	SLC1A5	19	47290716	cg15901722	0.675476	PCDH24	5	175974973
cg01343936	0.565329		13	24560026	cg15902297	0.665646	VOPP1	7	55614512
cg01345586	0.670199				cg15908708	0.660271667	CXCR1	2	219030830
cg01362605	0.520314667		22	46374836	cg15912800	0.679678	MIR196B	7	27209197
cg01366246	0.594569667	HES1	3	193852768	cg15947959	0.685004	ERGIC1	5	172261155
cg01372694	0.735633667		7	65878352	cg15952045	0.685502667	ITGA9	3	37848209
cg01380194	0.666059667				cg15959756	0.661592333	SPAST	2	32289083
cg01397521	0.655715333		2	232546381	cg15965241	0.66746	P2RX7	12	121570571
cg01400058	0.720430333	C9orf93	9	15553069	cg15978276	0.716593667	PIAS3	1	145575814
cg01403246	0.575166	MAFA	8	144512192	cg15989436	0.66115		5	150465875
cg01408932	0.719681	PPIL2	22	22020236	cg16000022	0.675493			
cg01417615	0.545926333	RAB3B	1	52456419	cg16007497	0.688777667	PPBPL1	4	74713356
cg01422881	0.679421667	RNF220	1	44873692	cg16024933	0.594346	TBL1XR1	3	176915599
cg01429933	0.654994	PLEKHG5	1	6556248	cg16026647	0.696081	MEIS3P1	17	15689920
cg01438737	0.743229333	SFRS6	20	42086396	cg16044674	0.694204333	MEA1	6	42981944
cg01445100	0.665102667	BANP	16	88103339	cg16047471	0.653252667	BHLHE41	12	26274563
cg01446164	0.73973		11	2421746	cg16061228	0.684769333	SCN4B	11	118014547
cg01447828	0.659786	PRX	19	40919465	cg16098780	0.645495333	PIIF	10	81107244
cg01465769	0.689211667	PRKCD	3	53226412	cg16114706	0.681473333	LOC400931	22	46509464
cg01470744	0.599485333	UCK1	9	134406824	cg16117472	0.550681	SLC2A10	20	45338396
cg01483119	0.670031333	FOXK2	17	80551543	cg16120422	0.744184333	CCDC42B	12	113590924
cg01493567	0.654131	ZNF84	12	133612436	cg16125725	0.636615333		15	70101302
cg01493617	0.726395	HGS	17	79650915	cg16148346	0.657097		11	58826475
cg01519765	0.667985333	ETV5	3	185828120	cg16153549	0.66352		2	3496821
cg01527777	0.664477333	PHOX2A	11	71956145	cg16171137	0.668965333	BAT3	6	31607634
cg01575096	0.548328333	SCLY	2	238969261	cg16201674	0.63572	SOX21	13	95364586
cg01576146	0.665503	FLYWCH2	16	2936204	cg16204618	0.699752667	ZNF320	19	53394626
cg01588444	0.664033		12	132924320	cg16206931	0.661504333	INPP5A	10	134503280
cg01588449	0.696639	D2HGDH	2	242682104	cg16222896	0.679581667			
cg01593552	0.701680667		13	22185680	cg16229180	0.694962667	GPX4	19	1103746
cg01604883	0.678799667		5	149674905	cg16234335	0.521883333	PCDHA2	5	140188119
cg01619509	0.691569333	PITPNM2	12	123476380	cg16236394	0.713539667	ZC3H6	2	113033084
cg01633149	0.710409333	ABCC1	16	16208885	cg16246545	0.527560667	PHGDH	1	120255941
cg01634153	0.609131667	MACF1	1	39571505	cg16246698	0.711971	PXMP2	12	133263907
cg01651593	0.703788	CDC20	1	43824241	cg16248798	0.660356667		1	247271246
cg01676996	0.68066	C6orf136	6	30619167	cg16257181	0.665938667	PTDSS2	11	448827
cg01688293	0.676960333	CHFR	12	133428407	cg16260126	0.540912		10	4892269
cg01692968	0.660196333		9	108005349	cg16268734	0.653845667	CS	12	56690194
cg01695406	0.553475	TMEM190	19	55889276	cg16279861	0.618163667		11	128531312
cg01733795	0.701136	SENP3	17	7465439	cg16313758	0.685839667	IQCE	7	2647039
cg01735357	0.550060667	HAR1B	20	61732467	cg16336872	0.681914333	AHRR	5	435267
cg01751802	0.600148	KANK2	19	11309639	cg16341592	0.743027	BOC	3	112931182
cg01753241	0.670906667	TMEM184A	7	1587459	cg16348254	0.664258667		10	43837992
cg01756756	0.665500333	C19orf12	19	30206951	cg16351010	0.677855667	SDHA	5	221513
cg01771416	0.660997	NFKBIZ	3	101567360	cg16370863	0.660071	S100Z	5	76210838
cg01785359	0.671206667	PLEKHA2	8	38757652	cg16377998	0.661813	L3MBTL4	18	6284498
cg01830765	0.734070667	CCT6A	7	56119079	cg16393156	0.660402333		1	179557438
cg01854842	0.727331	ADNP	20	49547693	cg16394018	0.66923		11	14927549
cg01856162	0.662263	ECEL1	2	233350940	cg16399751	0.662176667		2	73608838
cg01859460	0.705998	OAZ1	19	2273050	cg16405432	0.682401667		14	95973710
cg01862311	0.654667667		8	26308815	cg16427107	0.636661667	SNUPN	15	75912924
cg01866959	0.715276667	ATP11A	13	113495690	cg16427743	0.687407333	PLEKHG3	14	65210080
cg01868367	0.670960333		1	247372405	cg16445708	0.709033667	NME3	16	1822320
cg01876619	0.678981667	ECE2	3	183971267	cg16455894	0.657125333			
cg01879591	0.654400667		2	242954430	cg16458021	0.748381667	ZNF295	21	43430507
cg01886077	0.516263667	TMEM131	2	98612542	cg16488580	0.664433333		5	173097570
cg01888767	0.747680333		5	131832552	cg16493813	0.677990667	MTUS1	8	17521952
cg01904393	0.661739667		12	51815083	cg16498359	0.680001333	SLC29A4	7	5340232
cg01905589	0.544559333	RPS6KA1	1	26856408	cg16504798	0.664660333	MYO1F	19	8643540

Appendix A

cg01923337	0.660264	AMDH2	16	2571449	cg16508522	0.707509667	RASA2	3	141319423
cg01963056	0.695558	PUF60	8	144911482	cg16512867	0.692911	NCK1	3	136581189
cg01964852	0.666317	HOXA3	7	27146262	cg16554164	0.652063		1	32410678
cg01966117	0.659225333	STAB1	3	52528714	cg16570019	0.694807333	COL9A3	20	61448188
cg01999046	0.674935667				cg16571794	0.663129667	SORCS2	4	7648627
cg02003183	0.547116	CDC42BPB	14	103415882	cg16573386	0.729180333	CCNL2	1	1334508
cg02016753	0.672862667	ZNF672	1	249132838	cg16573782	0.689537667	CBLN1	16	49315536
cg02025034	0.676180333	HIP1R	12	123346710	cg16576049	0.59026	ADCK1	14	78266708
cg02052217	0.672241333	PILRA	7	99970869	cg16591304	0.586119	CASP7	10	115439377
cg02067430	0.673374667	SLC25A10	17	79679069	cg16639138	0.549840333	ZNHIT1	7	100861083
cg02083412	0.707111				cg16658737	0.692665667	ZNF529	19	37064572
cg02085294	0.653649	SLC1A4	2	65220148	cg16664617	0.669137	SLC27A1	19	17607007
cg02089380	0.715290333	ACVR1	2	158732891	cg16671360	0.668363667	DCI	16	2300569
cg02096396	0.656272333	FAM59B	2	26395556	cg16691477	0.658583	FAU	11	64889790
cg02098905	0.663876667		8	119759090	cg16722292	0.741943	KIN	10	7830180
cg02121447	0.658488333	SNX21	20	44461113	cg16736826	0.661372333	EDN2	1	41951512
cg02138358	0.621146667	MPO	17	56358318	cg16737517	0.520894333	ZBTB46	20	62406677
cg02153528	0.681403333	OBFC2B	12	56618648	cg16757281	0.689440333	C16orf13	16	685785
cg02155262	0.659524667	AGA	4	178363707	cg16818993	0.675626667	SFMBT2	10	7453471
cg02182114	0.717186667	LOC100134221	7	139876578	cg16869008	0.561195	DGAT1	8	145550615
cg02189843	0.657743667	S1PR5	19	10625775	cg16883533	0.554639333	PHB2	12	7079667
cg02211160	0.678722333	VARS	6	31761076	cg16887862	0.689580667	PUSL1	1	1243669
cg02234820	0.719988333	LPAR3	1	85358327	cg16943083	0.710098333	COLEC12	18	500817
cg02235659	0.659608	CLEC3B	3	45066580	cg16968115	0.748960667	WDTC1	1	27560829
cg02246992	0.657266	ATAD5	17	29158363	cg16979622	0.534615333	AZI1	17	79196755
cg02249648	0.74527	FBXL14	12	1703181	cg17058475	0.571038667	CPT1A	11	68607737
cg02254774	0.655973333	LOC441601	11	50257496	cg17095489	0.671403667		8	128264282
cg02254876	0.705341667		3	42922351	cg17103705	0.603209	KIAA0100	17	26972054
cg02258201	0.665803	HRCT1	9	35906148	cg17114866	0.655935		19	35222640
cg02268354	0.518261	RHOT2	16	718175	cg17153426	0.680739667	GGT5	22	24640743
cg02276314	0.658264	H6PD	1	9301104	cg17159890	0.654447	CALHM3	10	105233093
cg02283485	0.706057333	ZMIZ1	10	81003020	cg17163729	0.669164			
cg02286335	0.743143	ZWINT	10	58121068	cg17170278	0.678077667			
cg02330874	0.522874333		10	71339632	cg17172569	0.665774667		1	224711904
cg02336442	0.657054333		2	139658821	cg17206978	0.705023333	CENPA	2	27008819
cg02349096	0.580391667	RAB3D	19	11450198	cg17218495	0.736579	SMARCA4	19	11071743
cg02385710	0.636833333	ZNHIT2	11	64884962	cg17222234	0.630270333	CORO2B	15	68871405
cg02386575	0.559426333	QRICH1	3	49068057	cg17237320	0.688799333	LOC401052	3	10053000
cg02450725	0.709596333		17	75502873	cg17243044	0.660889	ZBTB12	6	31867945
cg02454053	0.700677333	ZNF592	15	85327774	cg17255450	0.744742667	EVC2	4	5710372
cg02455094	0.683467333	ONECUT2	18	55104078	cg17267676	0.563441667		19	1194783
cg02463426	0.718030667	SLC35E1	19	16683387	cg17323243	0.699820667	NDFIP1	5	141488153
cg02468154	0.717934	HSD17B12	11	43755634	cg17332338	0.590215333		2	38763354
cg02474109	0.661796667	MAD1L1	7	1912065	cg17360140	0.752357	C4orf29	4	128886135
cg02484850	0.641891	MYH9	22	36783942	cg17362661	0.668500667	AFF3	2	100210490
cg02513379	0.519348333	IL21R	16	27414281	cg17368874	0.606570667	RECQL4	8	145743346
cg02524531	0.683153		22	46041485	cg17418387	0.656633667	RCC2	1	17761802
cg02564536	0.578633667	ATE1	10	123687158	cg17425567	0.565068667		19	36422583
cg02566518	0.733815	FUT11	10	75531971	cg17456612	0.7032	ANP32A	15	69111367
cg02567756	0.721907333	ARMC8	3	137906401	cg17461336	0.714044667	CYP3A43	7	99441559
cg02578560	0.664949667	TTBK1	6	43243245	cg17488937	0.66838		2	11878864
cg02583796	0.658813667	ACBD3	1	226375174	cg17491846	0.706262667	KIAA0892	19	19431584
cg02603366	0.628978	RPS21	20	60962040	cg17492326	0.713359667	FAM91A1	8	124780746
cg02613601	0.722153333		13	32183063	cg17499324	0.659808333	FCGBP	19	40422963
cg02628823	0.544787333		4	141419587	cg17508639	0.655672667	PINK1	1	20959617
cg02629833	0.698802333		17	80349836	cg17512843	0.711275667	RNF34	12	121837458
cg02637654	0.673606333	SLC16A14	2	230934329	cg17517865	0.596732	KCND3	1	112532085
cg02691360	0.660030667	AMFR	16	56396075	cg17526060	0.701722	RNF25	2	219535845
cg02715602	0.541422667	SEMA6B	19	4544446	cg17527435	0.694968667	ARHGFE10	8	1900400
cg02717470	0.670489333		2	241559792	cg17550566	0.664664	TTYH3	7	2681436
cg02735762	0.700706333	TTF1	9	135280262	cg17580045	0.676370667	CCND2	12	4384890
cg02736280	0.658784333	WDR81	17	1633745	cg17588812	0.656975333	KIF1A	2	241689985
cg02753354	0.536056667	HMHA1	19	1074727	cg17626178	0.731077667	PARD3B	2	205410273

Appendix A

cg02759005	0.744570333	SELK	3	53925912	cg17629793	0.722549	ZNF574	19	42580205
cg02779230	0.616052667	TRNAU1AP	1	28879417	cg17630771	0.568296	YAP1	11	101981004
cg02818189	0.654521				cg17662034	0.742937667	RDH10	8	74207518
cg02821150	0.696272	DGCR9	22	19007132	cg17662369	0.657899	TBC1D9B	5	179333958
cg02827132	0.664736		16	1052529	cg17695841	0.670148	SLC35C1	11	45825485
cg02838932	0.528977667	CEP192	18	12991623	cg17714987	0.662795333	PIGY	4	89445475
cg02867162	0.637235	LMF1	16	1020843	cg17751366	0.679418		1	26551258
cg02888894	0.668087333	PRPF6	20	62660605	cg17764549	0.540349	PTPRN2	7	158287907
cg02892486	0.595886333		9	137028920	cg17779676	0.656459	CUEDC1	17	55982124
cg02893604	0.676111	PSMB11	14	23512805	cg17820890	0.677807	SFRP5	10	99531790
cg02902761	0.568776333	THRA	17	38219410	cg17822779	0.738597667		8	135174326
cg02942845	0.671242333	CRISPLD1	8	75897115	cg17832704	0.658466	SHBG	17	7516888
cg02947021	0.582262667	C14orf43	14	74227042	cg17895236	0.709696	NTN1	17	8924740
cg02973377	0.682591333	EPHX3	19	15342548	cg17901584	0.638216	DHCR24	1	55353706
cg02977954	0.612986	ECHDC2	1	53371368	cg17915189	0.666770667		7	43789728
cg02990302	0.655101667	C16orf80	16	58155189	cg17918280	0.716215333	PLCH2	1	2411378
cg02994863	0.740679	PGM1	1	64059297	cg17936236	0.65805	CREG1	1	167520305
cg03014241	0.682575333	SOC51	16	11348611	cg17941109	0.657676667	ABHD8	19	17407198
cg03014934	0.654497333	MIR548F5	13	36050901	cg17944940	0.73903	SLC25A19	17	73285520
cg03031660	0.747913333	MRPS7	17	73257791	cg17958516	0.662606667	ZMIZ1	10	80952606
cg03031868	0.525996667	ESD	13	47371523	cg17987384	0.688285333	FLJ40504	17	26634469
cg03040622	0.680546667	MIR589	7	5536937	cg17987601	0.568094	1-Sep	16	30389834
cg03043709	0.661475667	MAD1L1	7	2046920	cg18024053	0.666452333	ARHGEF10	8	1796568
cg03068319	0.671892333	FAM8A1	6	17600252	cg18030105	0.731875333	ZNF195	11	3396237
cg03068497	0.570857667	GARS	7	30635838	cg18030218	0.61034	PSMD8	19	38865513
cg03092918	0.739478	OAT	10	126107592	cg18034295	0.671345333	C22orf32	22	42475135
cg03156651	0.672317667				cg18036763	0.752493	PHF21B	22	45404910
cg03159660	0.682752667		11	2078197	cg18043267	0.680268667	C10orf125	10	135170752
cg03176386	0.721009		7	38370664	cg18045172	0.723308667	TAP2	6	32801648
cg03179542	0.659707667	IQCE	7	2647333	cg18064468	0.661003667		17	46570458
cg03188382	0.653233333	ALPP	2	233245886	cg18080903	0.683721333	RNF130	5	179499337
cg03194226	0.653777333	CLEC3B	3	45066832	cg18102469	0.719764667	CNBP	3	128902681
cg03206681	0.713289		3	193512175	cg18105735	0.662159	GUCY2D	17	7922470
cg03211327	0.675331				cg18108507	0.561326667	RRP9	3	51975897
cg03218192	0.654259	AP2B1	17	33914403	cg18153322	0.574687667		1	43534145
cg03222009	0.657004333	HS6ST1	2	129077232	cg18164305	0.662736667			
cg03225444	0.665351667		2	129494558	cg18167715	0.702023333	ARF1	1	228270190
cg03231960	0.739701667	DBP	19	49139050	cg18171204	0.692474			
cg03242458	0.661886	TRIM15	6	30135644	cg18181134	0.725822333		17	20467965
cg03243895	0.663534	SPTAN1	9	131314908	cg18181703	0.606706667	SOC53	17	76354621
cg03273509	0.728757667	VAPA	18	9913806	cg18208742	0.656169667	GFAP	17	42992872
cg03280245	0.658609333	RTN4RL2	11	57242445	cg18223939	0.671435667	NAGLU	17	40687822
cg03309938	0.674022667				cg18232130	0.664248333	ETHE1	19	44027824
cg03313945	0.654788333	ZIC5	13	100622158	cg18234089	0.667148333	SUPV3L1	10	70939902
cg03340026	0.683957		19	1403530	cg18244157	0.653410667	IGDCC3	15	65622965
cg03358636	0.671083	KIAA0226	3	197474006	cg18272543	0.655771333	TAX1BP1	7	27779524
cg03361810	0.727569333	AP3M2	8	42010162	cg18282011	0.622082667	CDK11B	1	1590447
cg03363633	0.727698667	TYROBP	19	36400317	cg18306406	0.675113	SYNGR2	17	76168382
cg03371770	0.582822333	CIAO1	2	96932024	cg18346576	0.661818667		2	27938071
cg03432814	0.669097333		12	117037220	cg18355925	0.719519	SRD5A3	4	56212207
cg03435608	0.729933667	CHMP2B	3	87276226	cg18384588	0.677383		22	46463747
cg03446427	0.667771667	FGFRL1	4	1019985	cg18392604	0.708092333	DLK1	14	101192386
cg03448816	0.662702333	FAM76A	1	28052370	cg18429793	0.672823667	C10orf137	10	127408055
cg03450842	0.675191333	ZMIZ1	10	80834947	cg18436491	0.680501333	LRRFIP1	2	238657876
cg03458344	0.541959333	C1orf129	1	170964477	cg18458509	0.669936667	SLC22A18AS	11	2920189
cg03471150	0.653682	IPO9	1	201797198	cg18465199	0.659315		6	37506304
cg03490766	0.667588667	SMOC2	6	169027019	cg18467978	0.676146667		19	47134906
cg03497652	0.551597333	ANKS3	16	4751569	cg18496317	0.726586	PRIC285	20	62205614
cg03497750	0.710421667	ZNF321	19	53445884	cg18526140	0.700726667		14	106444894
cg03517919	0.656912333	KIF1B	1	10398268	cg18549919	0.700724667		8	41757981
cg03521625	0.664014333		2	189475142	cg18555746	0.696109	IGSF22	11	18725854
cg03528784	0.660581333	SLC17A5	6	74363598	cg18565702	0.732643	TUBGCP3	13	113241891
cg03536846	0.641302333	DEPDC6	8	120886254	cg18584387	0.687393	CCND2	12	4384879

Appendix A

cg03539717	0.656308667	GADD45GIP1	19	13065086	cg18592538	0.669217667		2	176121788
cg03541338	0.658609333				cg18608055	0.560927667	SBNO2	19	1130866
cg03555914	0.670098	GLI2	2	121554810	cg18652923	0.755624	SH3RF3	2	109745344
cg03566814	0.653455	SMOC2	6	169019307	cg18704218	0.732188		5	33162408
cg03572680	0.688178333	ROBO4	11	124768554	cg18725681	0.708777333	FITM2	20	42940213
cg03573109	0.741914333		16	85849619	cg18751306	0.747131	MESDC1	15	81294292
cg03580568	0.703302667	EDNRA	4	148402748	cg18753480	0.679312	ATXN7L2	1	110026701
cg03597525	0.664246	SIGLEC10	19	51920435	cg18766912	0.739476667	UBE3A	15	25683909
cg03606774	0.658748	SLC5A6	2	27432830	cg18773260	0.701151333	HOXB7	17	46685292
cg03611151	0.683654	CNR2	1	24229581	cg18774318	0.559772	ZNF180	19	45004584
cg03625260	0.661932667		6	17102391	cg18787437	0.656878333		1	16499673
cg03631596	0.743662667	C13orf37	13	73301985	cg18807071	0.658132667	ZDBF2	2	207139408
cg03636183	0.604670333	F2RL3	19	17000585	cg18839746	0.647716333	COL9A3	20	61448150
cg03638940	0.544368667	9-Mar	12	58149384	cg18842310	0.679068333	MCF2L	13	113627484
cg03652525	0.657313333		3	72149324	cg18895476	0.670539667	TPST2	22	26962071
cg03663120	0.536447667	NUDT1	7	2284600	cg18914962	0.655249667	MAP3K5	6	136914796
cg03667621	0.701206333	MICALL2	7	1478715	cg18927077	0.655949667		14	103605358
cg03671660	0.737134333	PSMG1	21	40554907	cg18931633	0.657313			
cg03676485	0.657233333	LFNG	7	2563816	cg18951674	0.667986333		11	93641499
cg03685346	0.727035	EFNA5	5	107005999	cg18954051	0.655859	TMBIM4	12	66554863
cg03697316	0.650038667	LZTR1	22	21336741	cg19005168	0.670298333		13	27092836
cg03725309	0.701104	SARS	1	109757585	cg19032492	0.675141333	CHD3	17	7812665
cg03732936	0.595772333	DEGS1	1	224371029	cg19034708	0.743176667	PCM1	8	17780168
cg03745431	0.677006	C1orf95	1	226736713	cg19054219	0.697603	CELSR1	22	46805820
cg03748310	0.717103333	KIAA1875	8	145168450	cg19064302	0.684685	PRDM1	6	106536036
cg03758021	0.655712333		20	62208140	cg19074474	0.674278		5	105582234
cg03773636	0.667258333				cg19099833	0.705739667	CSRP1	1	201476290
cg03784628	0.713304333	KCNK4	11	64059937	cg19109538	0.657622			
cg03785076	0.662056667	SNED1	2	241936915	cg19128498	0.541333333	FBXL5	4	15656874
cg03800797	0.672225333	KRT81	12	52680245	cg19137806	0.535643333	INPP5A	10	134362170
cg03834947	0.545889667	MTFMT	15	65321822	cg19139589	0.697915667		2	25149093
cg03837313	0.654321667				cg19158176	0.667329	SSU72	1	1510364
cg03843951	0.705900333	DDC	7	50629634	cg19180056	0.666519667			
cg03849707	0.706801				cg19185384	0.604902333	LOC100190939	13	45914939
cg03875450	0.743009667	PTCH1	9	98269581	cg19203575	0.695877333	ZNF323	6	28322086
cg03888578	0.701107333	CCDC18	1	93646377	cg19219366	0.655204333	TRIM3	11	6495536
cg03903296	0.658899667				cg19233923	0.741086667	OTUB1	11	63753598
cg03907612	0.709762333				cg19240052	0.738953667	C22orf32	22	42475693
cg03909417	0.529854667	AMACR	5	34007809	cg19254163	0.713229333	GPR44	11	60623782
cg03916756	0.719853333	BAIAP2	17	79032487	cg19260606	0.682074			
cg03922648	0.660011	KLRC1	12	10607364	cg19263228	0.689200667	PRDM16	1	3191660
cg03928032	0.657756667	SLC38A10	17	79244651	cg19271753	0.512434333	TNRC18	7	5396704
cg03941745	0.703565333	DCTPP1	16	30442009	cg19335412	0.676745667	ACTA2	10	90694875
cg03945387	0.667140667	ZNF688	16	30584858	cg19350970	0.664715		2	238039148
cg03959945	0.699423667	PLXNA2	1	208326650	cg19372602	0.661034667		1	156116207
cg03960699	0.693762333	USP34	2	61697528	cg19373170	0.711595667	SFMBT1	3	53079105
cg03965340	0.563521333	TCEB2	16	2826984	cg19373649	0.683869667	CR1L	1	207817630
cg04004437	0.58522	ANKRD29	18	21242479	cg19375676	0.646617333	ADSS	1	244615173
cg04077069	0.621799	RHEB	7	151217159	cg19377421	0.673237			
cg04077199	0.655695	GGT1	22	24989175	cg19410789	0.61706	LOC440926	1	226249811
cg04086239	0.661603333	PRKCB	16	24067174	cg19418648	0.714109333	RGS14	5	176789979
cg04100397	0.669500667	MALL	2	110875551	cg19427338	0.662777		2	42566455
cg04106436	0.695259	CLEC4D	12	8665373	cg19439022	0.708900333	RBM39	20	34330364
cg04113056	0.695021333	PPP2R5E	14	64010798	cg19442470	0.683781333	CLU	8	27470225
cg04152326	0.681602333	TPO	2	1488314	cg19445684	0.663381333	FGF1	5	142077529
cg04158069	0.637196667		22	50982905	cg19448292	0.607537667	C20orf118	20	35504064
cg04161137	0.669399333	C1QC	1	22974180	cg19463703	0.714937667	ARHGAP9	12	57874033
cg04161365	0.692764667	DHRS13	17	27230393	cg19469357	0.665894333			
cg04165030	0.627155667	CLTB	5	175843293	cg19474833	0.669789667	HOXB2	17	46622899
cg04192862	0.712865333	BAMBI	10	28966472	cg19475108	0.65865			
cg04193083	0.724715667	NBR1	17	41323562	cg19494223	0.521622333	C19orf60	19	18700933
cg04200362	0.724922	RAB11FIP5	2	73341598	cg19497548	0.668073		1	26249349
cg04202267	0.684170667	LASS6	2	169431900	cg19509894	0.707602333	ELAVL1	19	8070923

Appendix A

cg04208403	0.727164	ZNF423	16	49525807	cg19536407	0.669573333	BAHCC1	17	79431331
cg04214946	0.689320333	CDKN1A	6	36651933	cg19539824	0.705133	C9orf69	9	139010884
cg04238983	0.59232	MIR612	11	65210600	cg19543348	0.658112667		7	63213975
cg04258457	0.698059	ERN1	17	62207789	cg19569340	0.681098333	RNF216	7	5821596
cg04265971	0.754490333	RNF4	4	2470698	cg19578751	0.655670333	EGFL8	6	32134011
cg04266460	0.625791667	SOC51	16	11348956	cg19583211	0.537839667	TBR1	2	162273185
cg04280480	0.695223	PPARA	22	46547117	cg19593285	0.657104333	E2F1	20	32267661
cg04286195	0.66317	BANP	16	88099702	cg19596020	0.672343	RHOV	15	41166388
cg04315264	0.717427667	RPL10A	6	35436252	cg19600442	0.654569	ALG1L2	3	129800495
cg04319462	0.719301333	ZBTB22	6	33284965	cg19617213	0.711971	HMHA1	19	1074926
cg04346637	0.691184333	PRDM16	1	3331154	cg19634252	0.676847667		4	38155144
cg04351258	0.704927667	ZNF764	16	30569043	cg19653417	0.658357667		12	132654924
cg04352550	0.665354667	PLBD2	12	113796401	cg19654061	0.661471667	ALPP	2	233243398
cg04357830	0.531792333	RUNX1	21	36261693	cg19656070	0.746627667	TMEM93	17	3571978
cg04362790	0.713591667		14	100223811	cg19682639	0.706360333	ATAD3C	1	1391102
cg04364194	0.746530333	ARL8A	1	202114085	cg19689415	0.656075667	SERTAD3	19	40949328
cg04366628	0.653875667		4	26290463	cg19695041	0.655984667	TACC1	8	38615330
cg04381888	0.665622333	HDAC7	12	48195549	cg19696103	0.664121333	ZCCHC10	5	132354130
cg04385144	0.701867	MIAT	22	27053419	cg19699893	0.687600333	LPPR2	19	11473353
cg04400653	0.690281667	MGC26597	6	7986462	cg19711782	0.692315	TMEM115	3	50395925
cg04437662	0.678981	VPS18	15	41186380	cg19727124	0.700953333	PLEKHG5	1	6553227
cg04439252	0.663290667	STAT3	17	40475282	cg19727381	0.669004667		8	855862
cg04481603	0.657079				cg19732599	0.714223667		12	123865592
cg04482794	0.705531333	ITPKB	1	226925181	cg19773045	0.66599	SLC27A3	1	153747600
cg04482923	0.517841333	TMUB2	17	42264046	cg19789392	0.655722	DAK	11	61099826
cg04488153	0.663780333	ABHD14B	3	52004130	cg19846491	0.675719667		17	79322924
cg04497512	0.694753333	CLIC6	21	36042224	cg19847945	0.687368667	IO-Mar	17	60783784
cg04509559	0.557011667	ERH	14	69864994	cg19852660	0.657804	KCNQ1	11	2846681
cg04517323	0.719284667	LAPTM4B	8	98788873	cg19859559	0.654773333		8	144041178
cg04526361	0.691284333		7	22007577	cg19868593	0.667079333	C6orf27	6	31744816
cg04546097	0.669363667	HGD	3	120401058	cg19889780	0.665004667	SPR	2	73114144
cg04549115	0.688623333	RWDD1	6	116892534	cg19906973	0.696079			
cg04553410	0.751416667	GBX1	7	150864885	cg19934294	0.666718333	MFSD10	4	2933421
cg04557677	0.688623333	JAK3	19	17959082	cg19949929	0.709387	MNT	17	2304007
cg04564646	0.586145667	MOV10	1	113217655	cg19951663	0.691348667	NEUROG1	5	134871634
cg04566018	0.697352667	PVR	19	45146967	cg19973604	0.696435	FBF1	17	73914282
cg04574090	0.671227667	KCNE3	11	74178749	cg20011051	0.712417667	C10orf78	10	105881722
cg04575058	0.667742	TSGA14	7	130080426	cg20012028	0.668174	UNC84B	22	39151999
cg04576025	0.728270333		5	1386550	cg20065216	0.662299		12	12625652
cg04600406	0.672476	UQCR	19	1605414	cg20080320	0.717470333	HERC2	15	28363860
cg04608779	0.666268		6	54878849	cg20083107	0.706366		8	12523148
cg04627240	0.701837	GSK3A	19	42746847	cg20086916	0.659958	FOXP1	7	4759814
cg04665180	0.723737333	FREQ	9	132934972	cg20095398	0.694703667		4	1571649
cg04686763	0.677973		11	64655583	cg20127035	0.529769	SETBP1	18	42260234
cg04687241	0.683616	DLGAP2	8	1616381	cg20137237	0.71105	DPY19L2P4	7	89749020
cg04759220	0.748516333	JMY	5	78532560	cg20152841	0.654344667	MOBK12A	19	2073190
cg04759439	0.688362667	ERC2	3	56502612	cg20153322	0.664947333	PXN	12	120703977
cg04761746	0.667733667	SCOC	4	141177756	cg20154947	0.657519333			
cg04794505	0.717329667	PBX3	9	128508709	cg20164601	0.668878			
cg04837231	0.733876667	PER3	1	7885239	cg20168412	0.715243667		3	13324132
cg04839835	0.656923		1	159881334	cg20179907	0.659906	PDE2A	11	72336908
cg04865110	0.599816333	JAKMIP1	4	6202558	cg20194454	0.687661	PSMB9	6	32821578
cg04872399	0.676418333	TNNT2	1	201346784	cg20246820	0.689205	H6PD	1	9324688
cg04910183	0.731504	CDC73	1	193090988	cg20263885	0.683786	C1orf130	1	24882207
cg04921578	0.655809333				cg20284508	0.698426333	CHCHD6	3	126668502
cg04944090	0.659917667		16	85635486	cg20292636	0.661118333	PLEKHA6	1	204328377
cg04949219	0.702782333	C11orf2	11	64863633	cg20297979	0.677131	LLGL2	17	73552110
cg04981611	0.678913333	KCNK12	2	47798477	cg20309353	0.668348	ALKBH5	17	18089940
cg04982488	0.662347667	AFAP1L2	10	116103698	cg20346503	0.668719667		2	128994402
cg04983687	0.525292667	ZFPM1	16	88558223	cg20362308	0.655775333	FAM18B2	17	15412141
cg04987302	0.510246		14	57476116	cg20365330	0.721021667	LOC619207	10	135269638
cg04988978	0.624416667	MPO	17	56359578	cg20366603	0.721542333	GPS2	17	7218821
cg04999421	0.685201333	PGS1	17	76412842	cg20367329	0.598336333	RASD1	17	17399941

Appendix A

cg05001598	0.685748333	KIAA0368	9	114246158	cg20385110	0.645009333	EIF3M	11	32605283
cg05008975	0.678124333	SCN4B	11	118022317	cg20387258	0.696995333		15	96960392
cg05029189	0.603981	ADCY5	3	123168386	cg20460166	0.663567	LRRC42	1	54423202
cg05031283	0.57191	POLR2I	19	36605406	cg20462855	0.738446667	HEY2	6	126070385
cg05032384	0.648417333	SBNO2	19	1138811	cg20471927	0.717478333	DIP2C	10	435953
cg05039463	0.567651333	TTC7B	14	91282525	cg20488673	0.667005		14	100497940
cg05052335	0.701132	RAPGEF6	5	130970657	cg20518446	0.674668	AHNAK	11	62315034
cg05064002	0.588302667	FCHO1	19	17858612	cg20563269	0.663386667		2	129104576
cg05090127	0.592803		16	1350789	cg20577878	0.651624333	PDP2	16	66914781
cg05105845	0.746817333	GCH1	14	55369781	cg20583432	0.663528333	HIVEP1	6	12045381
cg05133205	0.659236	PPT2	6	32121249	cg20595453	0.668296	VP552	6	33219392
cg05187833	0.633070333	NUP153	6	17706939	cg20611882	0.67574	GPR177	1	68696709
cg05198094	0.515113667	GAA	17	78075597	cg20620527	0.653431			
cg05226707	0.672714667	FABP3	1	31846699	cg20632873	0.741210667	ARHGEF2	1	155948126
cg05232802	0.685058667	NOM1	7	156742224	cg20640261	0.667884	MSH5	6	31707019
cg05235344	0.668047		6	11810356	cg20648385	0.659118667	ATG16L2	11	72525603
cg05252279	0.659639		10	81588174	cg20665341	0.668685333		4	152732024
cg05254511	0.674745333	TOMM22	22	39077689	cg20682376	0.657483667	TNFRSF13B	17	16875542
cg05256242	0.585375333		12	12876945	cg20691283	0.734149667	MEIS1	2	66661461
cg05275468	0.684846667	KAT2B	3	20082916	cg20691720	0.673329	OXER1	2	42989971
cg05291773	0.658567		21	37500642	cg20700731	0.659093	EDN2	1	41951480
cg05300158	0.741873	SETD7	4	140477727	cg20701901	0.655514667	MIR29C	1	207975360
cg05310613	0.538053333	HECA	6	139456890	cg20729301	0.714549		2	233368115
cg05315670	0.538996	BAD	11	64037304	cg20752878	0.655591333			
cg05316627	0.677370667		6	87861261	cg20777257	0.734995		5	173841038
cg05319880	0.529845	UHRF1BP1L	12	100536419	cg20781216	0.682317333	ZC3H3	8	144546287
cg05325390	0.746260333	COG7	16	23464715	cg20823695	0.670379667	PEX10	1	2345410
cg05328197	0.701535	MAPK4	18	48086680	cg20829347	0.674601	SOX1	13	112720927
cg05339414	0.681557333		12	125423730	cg20829379	0.685993667	SLC6A20	3	45838083
cg05340042	0.654041667	IRGM	5	150224636	cg20844851	0.656893	VGLL2	6	117591827
cg05363534	0.594296667	MIR612	11	65210645	cg20863756	0.542020667	C12orf53	12	6809797
cg05368971	0.654569	MORN1	1	2321364	cg20911165	0.707384333	MUC5B	11	1244397
cg05388545	0.663661333	CRLF3	17	29152968	cg20912770	0.666964333	ASCL2	11	2292428
cg05397010	0.654823667	PLA2G4F	15	42448259	cg20916427	0.66126	KATNAL2	18	44627440
cg05417615	0.7406	SLC10A7	4	147443478	cg20921874	0.574369333	ZNF700	19	12034925
cg05423018	0.670537				cg20951444	0.661952333	FNDCC5	1	33337112
cg05428781	0.701699	CNOT8	5	154237537	cg20960277	0.696394333		13	21835122
cg05438727	0.687688333	KCNQ1	11	2800047	cg20960611	0.682003333	VENTXP7	3	21445979
cg05446659	0.664391	PNPO	17	46026612	cg20962543	0.553971333	SYDE2	1	85666741
cg05462426	0.646648333	ARHGEF16	1	3370836	cg20966551	0.735796	MAST1	19	12949060
cg05480883	0.666964667	ITIH5	10	7614570	cg20968743	0.678882333	TSPAN18	11	44883965
cg05484788	0.513938	PHACTR3	20	58180357	cg21028463	0.746572667	MIR636	17	74733682
cg05518543	0.699994333	MAEA	4	1283165	cg21037943	0.677860333		8	854013
cg05536998	0.691023333		18	46500646	cg21049809	0.697736667	TTC7B	14	91283462
cg05563388	0.516762333	ARHGEF16	1	3371107	cg21050240	0.690599	LTBP4	19	41108423
cg05566961	0.609791333		3	128398685	cg21058391	0.655738		4	160027487
cg05575921	0.541361667	AHRR	5	373378	cg21068013	0.673205			
cg05576845	0.659332	PPP2R2B	5	146176618	cg21094949	0.628825	PCDHA2	5	140202574
cg05596135	0.673461667	TXNL4B	16	72128010	cg21108767	0.752357667	PILRB	7	99933721
cg05611648	0.65424	KLHL30	2	239050834	cg21120539	0.732383667			
cg05617801	0.665519		2	108008839	cg21139312	0.699138667			
cg05624300	0.724983		16	2517622	cg21142038	0.669931667	LGR6	1	202183444
cg05636175	0.654730667				cg21158633	0.678202			
cg05648472	0.657522	PRDM11	11	45232364	cg21160472	0.743466	ATF3	1	212782112
cg05664421	0.658927				cg21167563	0.660883667	CD58	1	117114916
cg05755017	0.698414667	HSPA8	11	122929522	cg21177096	0.582463667	MATN3	2	20212440
cg05772917	0.657982	MAPT	17	44027251	cg21184629	0.573321333	C9orf122	9	38621834
cg05787952	0.654766	JMJD8	16	734667	cg21187770	0.67803	KIF3C	2	26205876
cg05815411	0.711116	STRADA	17	61819127	cg21189321	0.691519333			
cg05820312	0.563015	TRAPPC9	8	141468672	cg21210537	0.639512333	MIR769	19	46522185
cg05871756	0.696482		5	173216171	cg21210789	0.525662	BAI1	8	143545390
cg05874561	0.683961667	SFRP2	4	154709828	cg21230425	0.658930333		7	64533826
cg05880490	0.713292333	TAAR6	6	132890978	cg21231302	0.668451333	RNF208	9	140116093

Appendix A

cg05886453	0.666213333	LRFN3	19	36436010	cg21240420	0.668983333	COL11A2	6	33148096
cg05892484	0.543540333	MAD1L1	7	2143507	cg21241424	0.656479333	CCDC33	15	74610683
cg05914439	0.611867333	MRM1	17	34958381	cg21267498	0.698819333			
cg05948157	0.710789333	PROX2	14	75330608	cg21273013	0.721018333	HNRNP	1	23670943
cg05950943	0.668855667	NTN3	16	2520971	cg21279955	0.65384	SLC27A3	1	153747551
cg05951221	0.571973333				cg21280392	0.669252	PHOSPHO1	17	47304116
cg05970992	0.6085	KIAA0232	4	6785456	cg21317428	0.676549667		17	1157267
cg05991820	0.522166333	ECHDC3	10	11785161	cg21335375	0.685815	CCL3	17	34417434
cg06002638	0.661376	ACACB	12	109592017	cg21338479	0.671162333	CUX1	7	101459330
cg06009267	0.680366333	EPPK1	8	144948764	cg21343895	0.685020667	ACSM4	12	7456343
cg06011334	0.642936667	KHDC1	6	73973128	cg21366673	0.664461	HLA-E	6	30459512
cg06014792	0.679977	ZBTB4	17	7382760	cg21369801	0.565813333	CSNK1D	17	80202961
cg06023161	0.658686667	SNORA38	6	31590572	cg21441256	0.656466667	MRPL42	12	93895324
cg06023901	0.712753667	BAG2	6	57037482	cg21442773	0.749198667	LYST	1	236029828
cg06026129	0.632596	LPFR3	19	822355	cg21446172	0.662675333	CAPN8	1	223745234
cg06033949	0.664406		6	155259167	cg21454760	0.684686	ARID1A	1	27087961
cg06038358	0.671570333	PROC	2	128176007	cg21460606	0.656217	TNXC	6	32055402
cg06059409	0.686193333	ARHGEF17	11	73036247	cg21461196	0.673075333	ELAC1	18	48494958
cg06059663	0.677687333	KIF26B	1	245319431	cg21549415	0.657151667	P4HB	17	79803535
cg06093920	0.662891667				cg21566642	0.558557667		2	233284661
cg06094607	0.664593667	ZC3H3	8	144561886	cg21587837	0.660311	NFKBIL1	6	31525894
cg06110458	0.730951333		16	85256305	cg21603679	0.659976333	CPXM1	20	2774729
cg06126421	0.679989333				cg21605986	0.746965667	KIAA1191	5	175788725
cg06135139	0.713856333	BHLHE23	20	61637856	cg21609024	0.559754667	LRP8	1	53795111
cg06151744	0.669619333	P2RX5	17	3595104	cg21616720	0.659881667	TNFRSF4	1	1150085
cg06155697	0.676797333	TUB	11	8120396	cg21631583	0.518137333	DVWA	3	15247517
cg06182359	0.668849	PFDN2	1	161088957	cg21702497	0.585281667	POLR3E	16	22309096
cg06184669	0.692209667	GABRR1	4	47034267	cg21705506	0.677974333	SLFN12L	17	33842181
cg06185789	0.661026		18	45057704	cg21717508	0.670891	SNX15	11	64793846
cg06202737	0.664336		2	128166279	cg21741284	0.687362333	CDKN1C	11	2908036
cg06228828	0.65735	DLEU2	13	50653048	cg21741689	0.610516		3	128327189
cg06241044	0.663755333	PAIP2B	2	71451310	cg21776286	0.699073667	PILRB	7	99964298
cg06251289	0.654588667	CMTM3	16	66639593	cg21808406	0.660954667	CD68	17	7482475
cg06269559	0.668545		2	15298785	cg21810733	0.680782333	HDAC11	3	13520763
cg06297391	0.693034	NHLRC4	16	616827	cg21821726	0.707867333	BRD1	22	50170866
cg06318676	0.653853		1	92077357	cg21840875	0.660921333	SMCR7L	22	39910631
cg06345909	0.660251667				cg21845957	0.722645333	CKB	14	103988428
cg06358566	0.526194667	RPIA	2	88991375	cg21901395	0.667183333		17	46604235
cg06364629	0.749649333	EIF2B4	2	27592939	cg21903625	0.701465333		3	12810386
cg06385118	0.714726333	LOC441046	4	144480544	cg21939092	0.667060667	TLE2	19	3029597
cg06395091	0.655871333	DYNLRB2	16	80574926	cg21952241	0.713636667	CHD6	20	40032459
cg06405765	0.731523333	PPF4C	16	30087414	cg21957156	0.527849667	SLC30A2	1	26372718
cg06407470	0.691289	SV2C	5	75379270	cg21957546	0.610949	SNAPC5	15	66790095
cg06434490	0.662972667		6	31494261	cg21981144	0.663377667		19	45248572
cg06442192	0.525527333	ZNF541	19	48059856	cg22014289	0.534681	SH3RF3	2	109952974
cg06451822	0.671193		17	36108713	cg22021022	0.674857667	NMNAT2	1	183275320
cg06467910	0.637599667	DEPDC7	11	33037542	cg22023046	0.607667	SNORD46	1	45241572
cg06477069	0.669106667	C20orf117	20	35493378	cg22035166	0.665701333	C7orf25	7	42951793
cg06496803	0.653291667	TBC1D14	4	6940915	cg22068629	0.682375333	LMNB2	19	2446633
cg06500161	0.515234333	ABCG1	21	43656587	cg22078638	0.633334667	CD55	1	207495363
cg06529756	0.66577	SV2C	5	75380303	cg22093805	0.561885		4	53588374
cg06555397	0.573863667	FOXJ3	1	42801023	cg22103736	0.689735667	AHRR	5	373887
cg06559575	0.669759667	IGFBP6	12	53490352	cg22118131	0.532772667	TBC1D2B	15	78369834
cg06582394	0.514184	CASR	3	121902622	cg22118465	0.667794333			
cg06584619	0.649274667	PANK2	20	3869523	cg22143352	0.6879	JDP2	14	75897841
cg06596307	0.535633	IGF1R	15	99405016	cg22164238	0.739481333	AMPD2	1	110162488
cg06601581	0.671434		16	85404654	cg22223655	0.742297333	RTTN	18	67872902
cg06617335	0.661710333	LOC91450	15	78286905	cg22264170	0.676324333		6	9722973
cg06618322	0.666952667	AKAP13	15	85923867	cg22290284	0.679638333	MST1	3	49726443
cg06618497	0.673673	C14orf72	14	102198604	cg22339313	0.663885333			
cg06619959	0.666905	IL17RE	3	9956506	cg22356527	0.706129333	AHRR	5	374425
cg06623197	0.660340333	MTMR3	22	30400763	cg22364126	0.567323		4	171964547
cg06633438	0.533181333	MLLT1	19	6272158	cg22365313	0.655723667	PAPSS1	4	108636320

Appendix A

cg06635946	0.704341333		22	46470016	cg22372096	0.677190667	ZFR2	19	3810782
cg06639874	0.510637667	MLPH	2	238417703	cg22376704	0.734164		19	4760016
cg06642177	0.747428667	SGK1	6	134496341	cg22418565	0.661909667	MMP11	22	24114690
cg06647068	0.663702	CHST11	12	104853274	cg22479226	0.659407	LARP1	5	154153028
cg06652085	0.655932	TCF19	6	31130261	cg22479299	0.674536	TMEM90B	20	24449704
cg06668829	0.667700333	EPPK1	8	144948781	cg22496859	0.699963667	BRF1	14	105766987
cg06669446	0.726349	ANKRD11	16	89475959	cg22512377	0.738829667	C22orf32	22	42475683
cg06683078	0.659715667	KRTAP12-1	21	46102242	cg22523050	0.695734667	CREB3L2	7	137564212
cg06690548	0.554976	SLC7A11	4	139162808	cg22529396	0.670069667	MYO7B	2	128293186
cg06693983	0.536359667	TMEM190	19	55889216	cg22534627	0.648707667	RASSF3	12	65004466
cg06699484	0.608692333		5	132166328	cg22544881	0.657203	FLJ43663	7	130712346
cg06703062	0.749644	MATN1	1	31191648	cg22546775	0.657342667	HELB	12	66697281
cg06704539	0.720383	ALKBH4	7	102105256	cg22594055	0.733176333		5	131832675
cg06710704	0.710049		3	159778596	cg22617878	0.593341333	ATP2B2	3	10417183
cg06710900	0.547225333	IQSEC1	3	13008958	cg22619910	0.675931	ACOT11	1	55096622
cg06713373	0.672959	SLC12A8	3	124837734	cg22626683	0.665524		1	172903051
cg06723904	0.654434667		10	94179809	cg22637865	0.655882	TTL1	22	43486524
cg06750635	0.597214	ATM	11	108093419	cg22663117	0.612499667	MRPL12	17	79670454
cg06752482	0.679256	C17orf64	17	58499816	cg22669566	0.689322667	NINJ2	12	679287
cg06753281	0.668073				cg22678402	0.664425667	FAM125A	19	17534351
cg06778713	0.709836333		15	98937852	cg22695339	0.653871	CHD3	17	7791630
cg06793581	0.657239333	C16orf87	16	46865077	cg22704520	0.747215	C2orf60	2	200820451
cg06817736	0.666520333	TBC1D1	4	38108809	cg22710329	0.56506	ANUBL1	10	46168446
cg06819923	0.663018333	ZIP2	16	21214508	cg22749051	0.656933667		17	80668944
cg06852395	0.682070333	EIF4A1	17	7478131	cg22788465	0.659172667	IL3	5	131398845
cg06865642	0.653833	BRD1	22	50174028	cg22804935	0.568926	TMEM175	4	926322
cg06868100	0.521731333	PRR15	7	29606349	cg22819767	0.660540667	C10orf47	10	11866910
cg06868413	0.740158	CTNNA2	2	80325984	cg22821300	0.663443333	RASGRF1	15	79296585
cg06889348	0.539082		1	114895466	cg22822890	0.665317333		4	68017791
cg06897927	0.608405	HMX3	10	124895500	cg22835157	0.590041		15	41218352
cg06915334	0.572224333		8	20161343	cg22835724	0.672241		2	205125510
cg06933752	0.735245667	FAU	11	64889788	cg22843797	0.721443667	TNFRSF10A	8	23082961
cg06942183	0.694854333	HOXB2	17	46622607	cg22871529	0.728124333	FNDC3A	13	49550815
cg06959340	0.662002333	JUB	14	23450238	cg22878489	0.594420667	B3GALT4	6	33245701
cg06973106	0.665083667				cg22881266	0.683011667	PTPRN2	7	157543942
cg06982805	0.672809333	FAR1	11	13690105	cg22891278	0.656308667			
cg06996423	0.669745333	CALHM3	10	105236233	cg22896354	0.667752667		17	21367200
cg07008193	0.677961667	CREB3L1	11	46318527	cg22904711	0.666692	KCNN4	19	44278628
cg07016356	0.735367333	CXCR1	2	219032435	cg22926842	0.665176667	CENPM	22	42336869
cg07016556	0.659046	BAHCC1	17	79393532	cg22933020	0.701935333	TBCA	5	77072365
cg07020987	0.658488	TSC2	16	2127147	cg22941668	0.565180333	MIR145	5	148810180
cg07022048	0.672085333	KRT7	12	52638592	cg22950163	0.674036333	MMP2	16	55513525
cg07022640	0.681319		2	132152916	cg22967516	0.726082333	IREB2	15	78730339
cg07032309	0.728900333	KLF10	8	103668104	cg23003085	0.712702667	PRDX5	11	64084321
cg07042289	0.673990667		10	90692494	cg23009387	0.66815	SMAD5	5	135468634
cg07054502	0.659529667	ANKRD11	16	89405478	cg23011817	0.708573	ZBTB9	6	33422290
cg07058086	0.691227	KIF13B	8	29120186	cg23024254	0.694671333		1	33433579
cg07060076	0.725156667	MMAB	12	110011452	cg23040782	0.664643667	DNAJC11	1	6762215
cg07061867	0.695339667	C3orf63	3	56716962	cg23074119	0.547765333			
cg07066369	0.672930333	CCND2	12	4384888	cg23076913	0.662489333	MYOM3	1	24435582
cg07082267	0.661387		16	85429035	cg23097139	0.657876333	SNUPN	15	75918757
cg07086847	0.610149	DGAT1	8	145550618	cg23108125	0.697922667	PRR7	5	176882586
cg07092212	0.670102667	DGKZ	11	46382544	cg23112563	0.694872667	LOC642006	7	57247871
cg07123227	0.743057333	WRN	8	30891112	cg23114262	0.654007667			
cg07132038	0.666506333				cg23121728	0.660750333		10	65732826
cg07141036	0.682547				cg23141825	0.714561	PYCR2	1	226111974
cg07143043	0.670636	BNIP3	10	133784298	cg23152216	0.653647667	KRAS	12	25405085
cg07175797	0.655302667		16	50317656	cg23152931	0.654268	MUC6	11	1027560
cg07180646	0.594285	TMEM51	1	15481541	cg23160522	0.655175333			
cg07181376	0.596332333	C16orf57	16	58035192	cg23176340	0.549099667	MIR7-2	15	89154196
cg07183637	0.544647667	STK19	6	31940692	cg23177839	0.657048667		5	176757703
cg07199524	0.670909	CRTC3	15	91074149	cg23190089	0.658737333	SLC22A18AS	11	2920209
cg07206725	0.616831	CDKL3	5	133702808	cg23220012	0.694802			

Appendix A

cg07219103	0.645189333		1	203242409	cg23245316	0.659109333	TSSC1	2	3260005
cg07226802	0.654137667	RHPN1	8	144457703	cg23247191	0.683794333	TPCN2	11	68815682
cg07232033	0.532402	AGAP1	2	236403031	cg23250494	0.607801	CHRNA3	15	78913474
cg07241479	0.656445667	PRRT1	6	32116225	cg23266743	0.666621	TMEM132D	12	130388198
cg07243161	0.686585333	MAP7	6	136871709	cg23267616	0.698751333	ODZ2	5	167231004
cg07258149	0.667726667	SH3RF3	2	109940114	cg23267944	0.524225	NCL	2	232329239
cg07280272	0.653615333	DLL4	15	41230536	cg23271099	0.6554	CUL9	6	43149472
cg07280371	0.720256333		6	10835069	cg23281075	0.637175333	PUM1	1	31538727
cg07296835	0.659713667	SFRP1	8	41134425	cg23281456	0.700845			
cg07298547	0.663375333		16	85585356	cg23304339	0.738314	ZNF527	19	37862045
cg07311024	0.519177333	GLIPR1L2	12	75785089	cg23317859	0.66024	EPB41L1	20	34742329
cg07314414	0.571809333	SAP130	2	128784923	cg23343073	0.588944333	KCNK7	11	65360449
cg07323055	0.738996333	FASTKD5	20	3140559	cg23363754	0.522119333	PRR15	7	29606082
cg07326074	0.640844667	C16orf59	16	2510388	cg23413567	0.702845333	FOXO1	13	41223175
cg07331616	0.562362667	MACROD1	11	63933573	cg23450972	0.670704667	MYO15A	17	18076288
cg07359545	0.557334333	GP1BB	22	19711327	cg23456692	0.649965333	EDEM1	3	5229596
cg07386061	0.544082667	NISCH	3	52492874	cg23478349	0.660919667	C15orf39	15	75492982
cg07417990	0.711336333		1	224051868	cg23494533	0.665275333		22	19712750
cg07418114	0.655229333	NTRK1	1	156836717	cg23517124	0.685405333	RACGAP1	12	50419441
cg07433154	0.717616333		11	2364496	cg23528705	0.660264333	UNCX	7	1275252
cg07438421	0.681599667	SERPINF1	17	1675235	cg23541257	0.535406	KCNN1	19	18096662
cg07511188	0.683910333	LETMD1	12	51442144	cg23573114	0.682884667	LOC647979	20	34634400
cg07526021	0.660573	MSH3	5	79950611	cg23587288	0.686210333	SLC30A3	2	27483067
cg07531182	0.674341667	XYLT1	16	17299114	cg23594584	0.683726333			
cg07534263	0.690273667		19	24184660	cg23599820	0.654637	KIAA0195	17	73456199
cg07560408	0.524425	SNORD119	20	2444631	cg23629166	0.664298333	PEX10	1	2345368
cg07565236	0.698529667	CDT1	16	88869998	cg23629561	0.725505	KRTAP5-7	11	71238382
cg07568430	0.660134333	CSF1	1	110452482	cg23635560	0.701686333		2	27473369
cg07572233	0.654417667		1	234689565	cg23647410	0.689841333		1	221051066
cg07597763	0.622532333	OSBP	11	59382877	cg23677882	0.695153333		7	157281167
cg07601932	0.711755667	NDUFA6	22	42486924	cg23681311	0.624912333	MAPK1	22	22221878
cg07619799	0.727030333	ACSL1	4	185747409	cg23688719	0.658669	LEPR	1	65935654
cg07643930	0.660469	ZNF598	16	2060265	cg23693683	0.682618333	PTH1R	3	46924113
cg07687398	0.681269667	PRKCD	3	53198666	cg23702568	0.653793667	WAPAL	10	88281502
cg07691609	0.662297	RNF126	19	662740	cg23706268	0.722539667	RNASEH2B	13	51484060
cg07695566	0.655899333	CYB561	17	61525112	cg23728060	0.705944667	GALNT5	2	158113471
cg07701241	0.698217	DGKZ	11	46367812	cg23758016	0.730201333	LLGL2	17	73521635
cg07707505	0.670834	TNFRSF8	1	12185435	cg23783768	0.730400333	RASSF4	10	45454944
cg07720856	0.740065333	PTMA	2	232572668	cg23792308	0.690859667			
cg07721276	0.655330333	ARRDC2	19	18119064	cg23850277	0.665110333	MMP11	22	24114770
cg07734253	0.757304667	CORO1A	16	30194717	cg23850283	0.657235667		2	114261369
cg07751331	0.680118667	KIF3C	2	26205865	cg23877401	0.693639333	KDM2B	12	122017247
cg07786668	0.749424667	ZFH3	16	73092391	cg23878494	0.662788667		11	63687937
cg07789552	0.539027333	STK40	1	36851195	cg23924306	0.664035333	C1orf177	1	55271927
cg07790826	0.652946667	FADD	11	70049435	cg23943801	0.709876667	RAB21	12	72149029
cg07797882	0.520566333		5	149698722	cg23962478	0.718057333	PIM3	22	50354086
cg07806715	0.608250667	NAPA	19	48018254	cg24016995	0.678163	PTPRE	10	129868149
cg07810884	0.654188333	TPM4	19	16178099	cg24024056	0.673005333		10	102415556
cg07813961	0.6108	C3orf71	3	48956311	cg24050047	0.676687		7	33767686
cg07820189	0.615354333		1	183149515	cg24078828	0.679426667	WWTR1	3	149375789
cg07824081	0.655528667	GDPD3	16	30124904	cg24079591	0.744506333	MARK3	14	103851511
cg07824177	0.661999667	ISCA1	9	88896457	cg24097153	0.697248333	RHOF	12	122231765
cg07824742	0.700567	DBH	9	136501784	cg24109980	0.691133667	SLC18A3	10	50818707
cg07831432	0.733803	CANX	5	179125659	cg24111025	0.532624667	TAP1	6	32819921
cg07836815	0.670046667	LIMD2	17	61778515	cg24119674	0.653687333	DLG5	10	79648146
cg07839457	0.548671333	NLRCS	16	57023022	cg24137216	0.664884333	PLEC1	8	145020381
cg07844518	0.67753		16	85387556	cg24146100	0.551348667	DOCK9	13	99737448
cg07857181	0.634520667	POLH	6	43543548	cg24163210	0.680642667		1	1713944
cg07872276	0.726001		10	82075236	cg24180066	0.571063333	ZNF787	19	56632673
cg07883124	0.690054667	MCF2L	13	113634042	cg241817758	0.679868	RADIL	7	4878701
cg07897701	0.562767	ABP1	7	150549370	cg24199599	0.704222	SOX18	20	62681243
cg07899956	0.514233333	FAM100B	17	74261249	cg24212377	0.665340333		11	73110989
cg07903677	0.673181	KCNA3	1	111218079	cg24256946	0.595242667	PTBP1	19	796991

Appendix A

cg07905568	0.661218333	ZNF764	16	30565391	cg24288581	0.67404	DSCAML1	11	117376331
cg07911523	0.733482	FBXO7	22	32871416	cg24304618	0.659012667	CYB5A	18	71959402
cg07917031	0.674781333	SIPA1L2	1	232571335	cg24309428	0.653386	NOX4	11	89224381
cg07919145	0.666866667	SYNGR1	22	39780408	cg24331354	0.648089	SLC6A6	3	14444067
cg07959491	0.629212333		1	204347150	cg24332710	0.662461	AKR1A1	1	46016173
cg07987587	0.744185333	NDUFA6	22	42486991	cg24341498	0.752026333	RXRA	9	137217390
cg08026791	0.723950333	FAM86A	16	5134964	cg24342283	0.700180667	CXCR5	11	118758603
cg08030282	0.687605667	F11R	1	160967582	cg24354901	0.735379333	KDM5B	1	202777590
cg08042975	0.733446	ERBB4	2	213402433	cg24375627	0.663426	S100A6	1	153509284
cg08053103	0.656381667		19	5799467	cg24395386	0.670686667	PRDM16	1	3328593
cg08079166	0.661003333	MAP2K5	15	68083412	cg24408896	0.671237	SLC27A3	1	153747658
cg08089128	0.685854333	POM121	7	72391675	cg24447788	0.520539333		19	795310
cg08106661	0.553969333	TAF11	6	34855635	cg24475210	0.749335	MRFPAP1	4	6642433
cg08128789	0.671223333	LRRC8B	1	89989925	cg24504831	0.653658		8	145854306
cg08131547	0.726116	ZNF121	19	9691569	cg24526499	0.703441667	EFNA1	1	155100382
cg08134053	0.688738667	MFS2D8	2	24232856	cg24540603	0.712690667		6	31547917
cg08134342	0.585103667	LOC338799	12	122239035	cg24550676	0.653880333	DND1	5	140053903
cg08174462	0.719502	CNPY2	12	56709715	cg24561572	0.739652667	FMNL1	17	43298813
cg08184047	0.68586	ANAPC11	17	79849980	cg24568905	0.703375333	GFR2	8	21614417
cg08193363	0.530293333	FRY	13	32605254	cg24578966	0.694996333	AGPAT9	4	84457523
cg08201152	0.611941667		19	1875879	cg24608684	0.749677333		6	130686865
cg08241117	0.667649667	SLC2A11	22	24199628	cg24609304	0.699001	NLRC4	2	32483287
cg08248134	0.659684667	MCF2L	13	113659597	cg24610274	0.532461667	RAG1AP1	1	155107802
cg08248751	0.667449		6	44043192	cg24613083	0.568232333	TBX4	17	59539715
cg08252533	0.587559	UBE2F	2	238875575	cg24621042	0.660174	SERPINA1	14	94857275
cg08253970	0.701860333	C12orf66	12	64615825	cg24654547	0.746454333	DUS2L	16	68057165
cg08272151	0.626746333	SYNPO	5	150036071	cg24681499	0.705842333	FOXJ3	1	42706997
cg08273501	0.673667				cg24694018	0.653460333	POLR3GL	1	145457621
cg08276755	0.664303		20	30196714	cg24707432	0.660551	ZNF142	2	219513555
cg08300346	0.707186333	ZBTB16	11	113932106	cg24727122	0.698419333	OSM	22	30662972
cg08309687	0.533441667		21	35320596	cg24748448	0.688492667	TRPV1	17	3493352
cg08318582	0.563034667		11	109920663	cg24756528	0.750371667	CORO7	16	4466888
cg08331313	0.657594	SPARC	5	151066460	cg24769167	0.704677333	STK11	19	1205715
cg08346159	0.70285	NPY5R	4	164265012	cg24781737	0.687648	KLC2	11	66034681
cg08353938	0.655966333	4-Mar	2	217227472	cg24790600	0.679876333		10	123368753
cg08354681	0.667324	CDYL	6	4773278	cg24795903	0.675503667		17	36999666
cg08355863	0.665406667		6	33587496	cg24798853	0.706296		11	58866721
cg08362785	0.656253667	MKL1	22	40814878	cg24819738	0.738245	NR2C2AP	19	19314415
cg08364767	0.688124667	GPR176	15	40212176	cg24842354	0.658011667	PRKCZ	1	2004057
cg08371947	0.655886667	C4orf23	4	8477793	cg24845274	0.676258667	CHRM2	7	136555697
cg08409562	0.657535333	C6orf27	6	31737885	cg24847046	0.673112667			
cg08426951	0.674862667	CCDC61	19	46521569	cg24847937	0.686496	CARS2	13	111294884
cg08427928	0.684352	FKBP9	7	32997111	cg24848615	0.672512333	NFIC	19	3368396
cg08458912	0.691803667	MAP7D1	1	36644235	cg24860819	0.722354333	CHRN2	1	154544088
cg08492619	0.667551333	SLITRK5	13	88324193	cg24870676	0.705207667	FAM22F	9	97090861
cg08497766	0.648158333	IL15RA	10	6019877	cg24873093	0.658572667		1	45991060
cg08499057	0.714553667	ABCF2	7	150909642	cg24886770	0.687985	ZNF43	19	22018538
cg08503002	0.711960667	ABCF3	3	183903768	cg24896109	0.669711333	CYP51A1	7	91764007
cg08516217	0.658794667		19	38887799	cg24959663	0.705140333	ANKRD33B	5	10578618
cg08521010	0.599752333	USP13	3	179370821	cg24963001	0.697595			
cg08526825	0.693295667	SRRM2	16	2802229	cg24977276	0.533069333	GTF2I	7	74105270
cg08550882	0.741608333	MICALCL	11	12308354	cg24979495	0.667125333	TRIM50	7	72738631
cg08567142	0.665539667	HK1	10	71087924	cg24990400	0.669228667		8	102006838
cg08579753	0.756726333				cg25002152	0.663187667		10	74036833
cg08617970	0.654606	VARS2	6	30881112	cg25006998	0.674795		11	119897638
cg08643824	0.667023	LPXN	11	58294523	cg25033990	0.678883667	ACCN1	17	32484014
cg08667148	0.709160333	TNFAIP3	6	138188411	cg25051278	0.664374667	CFLAR	2	201980838
cg08676905	0.709736	IL15RA	10	6019609	cg25072436	0.745134	LBH	2	30454560
cg08678851	0.664341		5	91710951	cg25078026	0.66855		6	28829931
cg08701429	0.579884667		17	15394553	cg25082487	0.687426			
cg08718459	0.690814333		16	88355883	cg25094927	0.568248667	RANGAP1	22	41682102
cg08726522	0.667477667	ST5	11	8739587	cg25139493	0.681355	BMP8A	1	39957400
cg08739576	0.741742	AEBP1	7	44144360	cg25142954	0.733672	CHAC1	15	41245554

Appendix A

cg08764167	0.730321667	BTRC	10	103113933	cg25150878	0.655377667	HDAC11	3	13520921
cg08786207	0.680919667	ANK1	8	41559303	cg25182165	0.675953667	AKAP1	17	55191642
cg08787083	0.654248333	PTPRE	10	129868324	cg25183890	0.650507333	PDP1	8	94929275
cg08796240	0.536818667	VAC14	16	70733832	cg25189564	0.693443667			
cg08803263	0.659902667		16	80966302	cg25207224	0.656095333	HMGA1	6	34211189
cg08806699	0.591073	RBP7	1	10057970	cg25212453	0.669659			
cg08846112	0.707187667	SNTB1	8	121598484	cg25236971	0.623606667	FAM184B	4	17783059
cg08853494	0.622958667	RCHY1	4	76439657	cg25292468	0.676937333	NRP2	2	206550511
cg08859206	0.543583333	SCP2	1	53392774	cg25292882	0.687317		15	39431467
cg08867667	0.672673		10	95605214	cg25294185	0.723356	RNASEH2C	11	65487814
cg08867767	0.678514333	GIA9	1	39342651	cg25301142	0.709124333	HMGXB4	22	35653619
cg08869244	0.680279333	NDUFV1	11	67373841	cg25308231	0.729821667	TBCEL	11	120894785
cg08888487	0.715238667		5	1124759	cg25318364	0.707059667			
cg08922729	0.657635333		1	21913557	cg25320169	0.680424333	NRP2	2	206641630
cg08931196	0.665775667	RNF26	11	119205177	cg25347941	0.675095333	TTL1	22	43486472
cg08962087	0.707743333		13	29107120	cg25371036	0.695055333	AMOTL1	11	94500749
cg08965276	0.586555667	SOX8	16	1030586	cg25396690	0.704403667	CMPK2	2	7006004
cg08965527	0.721631333	HSDL1	16	84178213	cg25404033	0.714604333		7	108297992
cg08973950	0.688013	C7orf50	7	1083309	cg25420477	0.538918333		2	70319121
cg09000178	0.744905	CBFB	16	67063319	cg25469474	0.700133667	TNIP2	4	2749501
cg09122913	0.646666	FAF2	5	175875427	cg25487903	0.547677	DNM2	19	10828950
cg09133706	0.689543667	SIRT5	6	13588339	cg25497530	0.549788667	PTPRN2	7	158059944
cg09139721	0.659568		17	77509737	cg25498327	0.736744333	SLC30A5	5	68389852
cg09150006	0.737219		6	158652830	cg25503804	0.666494			
cg09152259	0.587379667		2	128156114	cg25508605	0.714239	SLCO3A1	15	92706125
cg09167414	0.671985333		1	16076206	cg25514992	0.672908		13	60014335
cg09186748	0.691621667		17	73577594	cg25535418	0.669959333	PLAGL1	6	144386474
cg09202659	0.653607667	JARID2	6	15380565	cg25551358	0.665381667		5	139079288
cg09238332	0.707253333		16	89070903	cg25555336	0.653418	MSI1	12	120806880
cg09238957	0.753676333	ORC6L	16	46723420	cg25557306	0.663302		8	98305878
cg09281539	0.745911	RALGAPA2	20	20693126	cg25575688	0.655885667	SOX12	20	305917
cg09293614	0.655220667	MUM1	19	1378043	cg25589371	0.735344333			
cg09308580	0.654059		2	43405947	cg25597115	0.727224667	BRP44	1	167906390
cg09325164	0.673974667	TMEM39B	1	32537942	cg25607383	0.675041667	DDR1	6	30853569
cg09349128	0.589686		22	50327986	cg25611931	0.641667667	C2CD2L	11	118978508
cg09353063	0.709242333	OXTR	3	8811092	cg25614222	0.676113	ZBTB44	11	130184756
cg09373298	0.654117667	CHCHD6	3	126613533	cg25684999	0.680073333	PTF1A	10	23480803
cg09384035	0.678549667		8	104131768	cg25703784	0.607662667	WDR46	6	33255181
cg09414330	0.664173		17	21409603	cg25707994	0.751175	DNAJB6	7	157129685
cg09423312	0.575426333	C7orf50	7	1163549	cg25723585	0.669900333	RNF165	18	43913649
cg09445967	0.659203	ZP1	11	60634850	cg25767906	0.599047667	SCP2	1	53392781
cg09469355	0.594042667	SKI	1	2161886	cg25769469	0.617053	PTCD2	5	71643841
cg09473207	0.655962667	RBM16	6	155079453	cg25781121	0.742009667	ZNF589	3	48282641
cg09477124	0.668020333	ATP2B2	3	10379281	cg25783241	0.729836333	AIRE	21	45713607
cg09514055	0.746190333	HYI	1	43919573	cg25842285	0.662561	CNNM1	10	101089972
cg09636302	0.528693333	HAL	12	96389483	cg25858682	0.730611667	CCDC72	3	48481793
cg09643587	0.647440333	CD47	3	107809710	cg25869317	0.729677667	CHSY1	15	101792241
cg09646173	0.659306667	PDE6A	5	149317669	cg25889711	0.538412333		10	102473609
cg09682727	0.670043		5	139953966	cg25894160	0.665406667	C20orf3	20	24974513
cg09717545	0.564746667	KISS1R	19	917707	cg25910466	0.716378667	XRN2	20	21283998
cg09750643	0.694311	HCCA2	11	1718086	cg25941354	0.719582667			
cg09754110	0.683782667	CASZ1	1	10714160	cg25949513	0.664476333	CAMK1	3	9811735
cg09761846	0.665661667	P2RX3	11	57117162	cg25962774	0.634268333	F12	5	176831364
cg09771049	0.744377	KPNA2	17	66031798	cg25974903	0.642176333	RAVER2	1	65211057
cg09776718	0.676567667	PLEKHF1	19	30164820	cg25979108	0.581422	C19orf60	19	18700553
cg09780955	0.585871667	SORBS1	10	97321190	cg25984791	0.583313667	DDX10	11	108535797
cg09794680	0.605710333	CPNE6	14	24540452	cg25998745	0.674336		8	142028625
cg09813817	0.675245	ZNF692	1	249152821	cg26033520	0.665323		10	74004071
cg09820716	0.702653333		8	8580928	cg26049379	0.654444333	TMEM82	1	16071339
cg09829013	0.656308333				cg26050159	0.655862667	ADAM11	17	42837242
cg09837656	0.658737667	PLEKHO2	15	65149245	cg26050838	0.730372667	CASP2	7	142985210
cg09857761	0.706868	MSLN	16	815553	cg26072705	0.719904	C17orf51	17	21454773
cg09888721	0.712841333	ETS2	21	40178652	cg26090072	0.673228	RTKN	2	74669387

Appendix A

cg09900521	0.657657333	LRRC43	12	122677357	cg26147845	0.719298	EP400	12	132433838
cg09915769	0.715308667	KIAA1543	19	7660977	cg26151531	0.738516667	GNB1L	22	19842652
cg09916783	0.656642333	POMC	2	25391911	cg26158150	0.667751667			
cg09955705	0.666446	RBPMS	8	30253570	cg26159905	0.653839	ASB10	7	150884914
cg09969011	0.634257667	LOC440356	16	29874179	cg26161820	0.655208	STARD3	17	37792777
cg09977800	0.586113667	FOXK2	17	80477464	cg26163374	0.561814333	COG1	17	71188695
cg10009224	0.678267333	PCGF3	4	742497	cg26185554	0.677557667	MYO1C	17	1391406
cg10010533	0.687096333	FAM181B	11	82445170	cg26212924	0.587772667	CYC1	8	145150439
cg10017105	0.527638333	SLC44A4	6	31832428	cg26224725	0.701133333	WBSCR16	7	74489898
cg10021786	0.674504667	GPT2	16	46964129	cg26258108	0.737992333	NCL	2	232329189
cg10041390	0.748980333	PTEN	10	89623018	cg26259758	0.681585		13	100609125
cg10064322	0.518238	GIPC1	19	14591203	cg26260369	0.742253667	VTI1B	14	68141723
cg10073091	0.594644333	DHCR24	1	55352301	cg26262049	0.660684			
cg10084993	0.661886333	SLC9A3R2	16	2077647	cg26278858	0.733838	BMF	15	40401256
cg10097464	0.668032	ANKDD1A	15	65204370	cg26289871	0.667905667	LPIN3	20	39969385
cg10118167	0.658444333	PLXNC1	12	94676545	cg26292895	0.666076667	RBM20	10	112444795
cg10122474	0.748985	GALC	14	88459471	cg26295057	0.667538	GDNF	5	37839829
cg10122932	0.743925	MCM7	7	99698990	cg26309134	0.653544333	ZNF542	19	56879571
cg10130374	0.713868667	ARHGEF7	13	111935412	cg26325335	0.697700667	CACNA2D2	3	50402333
cg10139614	0.711374333	APRT	16	88878480	cg26325791	0.579295	MAST3	19	18234711
cg10142520	0.659307333	EHBP1L1	11	65344604	cg26335760	0.663561333	RAB6B	3	133606960
cg10196558	0.659031333		1	155043921	cg26336164	0.731126667	FEM1A	19	4791680
cg10241347	0.733604	LOC399815	10	124639260	cg26343512	0.702826333		14	77413728
cg10246774	0.675316667	KIF1B	1	10386268	cg26353844	0.705224		1	95133907
cg10250355	0.690009667	CYB5R3	22	43045158	cg26372517	0.525049333	TFAP2E	1	36039159
cg10251328	0.684977333	IGF2BP1	17	47074713	cg26376809	0.695345667	USP13	3	179370824
cg10294200	0.696953	NAA25	12	112546905	cg26445541	0.680480667	EP400	12	132498837
cg10295360	0.625370333	TRIM47	17	73874699	cg26453130	0.677340333	NCRNA00171	6	30010907
cg10307345	0.522359333	PTPN5	11	18771567	cg26462404	0.539632667	OAZ1	19	2270807
cg10314752	0.664049333	PPP4R4	14	94640828	cg26466234	0.690022			
cg10315945	0.670463		1	17236551	cg26467725	0.591103333	SLC03A1	15	92647041
cg10341991	0.746112	CDC25A	3	48229869	cg26470101	0.576706333		2	173099597
cg10381440	0.689159667	HOXB2	17	46622715	cg26470501	0.632941667	BCL3	19	45252955
cg10400937	0.672448	ZNF790	19	37329330	cg26501027	0.724917667	RAET1K	6	150326334
cg10420527	0.670986	LRP5	11	68138505	cg26516004	0.668258	CYP1A1	15	75019376
cg10423910	0.570062	SLC12A2	5	127420075	cg26538931	0.703334		16	2073293
cg10433678	0.707186667	EPPK1	8	144943023	cg26545346	0.691856	ESRP1	8	95682096
cg10433812	0.696572667		11	10952532	cg26554302	0.660406333		15	22467326
cg10440877	0.663063		2	208378475	cg26561196	0.714794667	SLC7A6OS	16	68344680
cg10450108	0.602901667	PCDHA6	5	140208893	cg26563141	0.681552667	RGPD2	2	88124876
cg10460130	0.737604333	DTYMK	2	242625978	cg26570218	0.744109333	DNAJC27	2	25195142
cg10462093	0.640523667	SCARB1	12	125348448	cg26578621	0.614604333	ARHGAP20	11	110583377
cg10488476	0.517787333	ITGA3	17	48133676	cg26599209	0.713242	CYFIP2	5	156692972
cg10491242	0.721076333	PNKP	19	50366052	cg26608221	0.715116	MOGS	2	74692293
cg10512203	0.679779667				cg26628811	0.654803			
cg10520226	0.694262				cg26636312	0.701918	MUC6	11	1023573
cg10549088	0.678536667		3	64277154	cg26649251	0.747049	ZNF224	19	44598564
cg10552847	0.729476333	HMGCR	5	74632966	cg26662201	0.701246667	BMF	15	40398307
cg10583632	0.640624333	ZNF624	17	16557199	cg26664457	0.684460333	ZNF652	17	47394423
cg10583893	0.655727667	ZNF808	19	53039743	cg26672098	0.681851		2	42326070
cg10588135	0.666785333	BCAS3	17	59329903	cg26675329	0.697542667	IL18RAP	2	103033753
cg10589813	0.665968667		20	48809978	cg26705583	0.609110333	SGTB	5	65017946
cg10602180	0.655058667	CLDN18	3	137729133	cg26725153	0.672550667			
cg10611580	0.667836	9-Sep	17	75319942	cg26740481	0.732271333	NDN	15	23931011
cg10632229	0.534844	H3F3A	1	226250431	cg26761618	0.667015	P2RY2	11	72928926
cg10636246	0.552301333	AIM2	1	159046973	cg26775538	0.666916667	LPPR3	19	815090
cg10671907	0.650817	CD164	6	109703593	cg26775914	0.740534	DHX34	19	47852455
cg10672726	0.666476333	AP2S1	19	47353150	cg26781886	0.690817333	EMB	5	49737336
cg10675120	0.553011	FUT11	10	75532504	cg26789698	0.673565333		16	11480327
cg10695831	0.713797		1	98519324	cg26811927	0.657997333	EPHA8	1	22904862
cg10702366	0.523693667	FGGY	1	60070383	cg26815395	0.659838			
cg10706966	0.714392333	CSTF3	11	33183108	cg26821960	0.680716333		11	74809555
cg10724442	0.723131667	SPTBN4	19	40971809	cg26825412	0.657094	SOX18	20	62681428

Appendix A

cg10726394	0.735154667	ARID1B	6	157098467	cg26841647	0.662309667	BAIAP2	17	79010968
cg10775230	0.712761	TMIE	3	46742491	cg26848724	0.658473	ALOX5AP	13	31326405
cg10781828	0.668506				cg26864395	0.677945667			
cg10789151	0.681289667	EDEM2	20	33735143	cg26873439	0.729836333	PTPN1	20	49126821
cg10798225	0.658688667	PTDSS2	11	448822	cg26874229	0.576364		2	105853672
cg10800833	0.672697	HOPX	4	57521657	cg26904049	0.681005333	DUSP27	1	167091094
cg10802414	0.696133	HCN3	1	155252776	cg26904702	0.681010333	SLC27A3	1	153746899
cg10845147	0.696890333		5	172149624	cg26937868	0.660876667		1	26202601
cg10849092	0.714742333		15	40574697	cg26955337	0.663163	FHOD1	16	67264367
cg10859726	0.553908		21	15383750	cg26963277	0.606624	KCNQ1OT1	11	2722407
cg10897459	0.663511333		7	73237042	cg26972881	0.692600333		4	161403006
cg10902738	0.65343				cg26990733	0.654794	MICA	6	31371465
cg10935662	0.510632333		3	72307754	cg26992566	0.557999333	COX4NB	16	85814166
cg10950593	0.679721667	CYTH3	7	6212775	cg26996201	0.740884333	PAK1	11	77122864
cg10961055	0.692728333	ARL3	10	104473654	cg27000944	0.668028667	C17orf95	17	74721842
cg10967866	0.570512	INPP5A	10	134362164	cg27007358	0.717789667	SP4	7	21467723
cg11011533	0.666495333				cg27030541	0.682432667	REXO1L2P	8	86568090
cg11016194	0.61349		1	42501599	cg27086020	0.637931	ADRB2	5	148206096
cg11024682	0.572190667	SREBF1	17	17730094	cg27101899	0.665305		17	19368130
cg11053496	0.624982	TXNDC16	14	53019545	cg27118809	0.620513667	TTC38	22	46663960
cg11059760	0.673924667	SF3B2	11	65836288	cg27175509	0.674467		2	149891051
cg11067964	0.658618667		1	36584112	cg27177841	0.672623	VASH2	1	213125950
cg11075509	0.656410333	PMF1	1	156182236	cg27184867	0.664187667			
cg11128285	0.705737333	C21orf15	21	15221404	cg27193080	0.660742	IFT122	3	129160612
cg11145461	0.665959	CSRP1	1	201475451	cg27215475	0.712552	SDK1	7	4146971
cg11162385	0.746846333	NANP	20	25604740	cg27241845	0.665406333		2	233250370
cg11171811	0.523810333	LTBP3	11	65326725	cg27246571	0.515899333	HAL	12	96389588
cg11173422	0.675175		1	41832478	cg27258878	0.669644667	RBPMS	8	30407398
cg11183227	0.530767667	MAN2A2	15	91455407	cg27261412	0.656214	WIZ	19	15550742
cg11199399	0.589542667	LTBP1	2	33172442	cg27289478	0.676576	RNF4	4	2470262
cg11220565	0.675875		20	47934802	cg27353073	0.662969333	PRKRIR	11	76092311
cg11220663	0.744717333	ADD2	2	70994863	cg27383975	0.559126333	NR2F6	19	17356182
cg11229390	0.722853	PPT2	6	32122082	cg27400746	0.658824667	SPIRE2	16	89904261
cg11231349	0.656633				cg27424206	0.693418		6	13872211
cg11252792	0.521789333		3	125932053	cg27449973	0.655551			
cg11287851	0.735725333	MCAM	11	119187772	cg27464065	0.692903	APOLD1	12	12940610
cg11289281	0.579344	GLYR1	16	4897921	cg27477373	0.661452	ZNF542	19	56879645
cg11295113	0.675821	FOLR2	11	71927938	cg27483469	0.656147	DOPEY2	21	37615000
cg11296937	0.735126667	PRDX5	11	64084966	cg27527993	0.671044667		15	74417266
cg11299854	0.699941	CCNI2	5	132083184	cg27537125	0.515544333		1	25349681
cg11330104	0.717784333	ZNF808	19	53030705	cg27553231	0.688505	UNC13A	19	17797533
cg11365360	0.655031667	EPHB3	3	184279509	cg27566403	0.719915	CYTH1	17	76719612
cg11368628	0.705864667	SYT8	11	1856183	cg27573593	0.706589667	PIK3CD	1	9778887
cg11372422	0.692076667	UGT8	4	115519985	cg27575692	0.660364333		22	23802502
cg11385013	0.714366	LOC729991-M	19	19256748	cg27582168	0.692291667			
cg11416290	0.655369333		4	111532410	cg27585074	0.620984	CSNK1G2	19	1947960
cg11429664	0.729802667	BLM	15	91260578	cg27586255	0.691324333			
cg11431585	0.703777333	RLN3	19	14138834	cg27586682	0.693859	PDIK1L	1	26437624
cg11433608	0.669285333	B9D1	17	19246574	cg27619163	0.748981333	ALOX12B	17	7982806
cg11470337	0.656656	NPC1L1	7	44579271	cg27624327	0.668204667	SLC9A3	5	510254
cg11474864	0.5877		18	60263588	cg27629977	0.662136667	CTNNA2	2	80531633
cg11488020	0.699991	KIAA1688	8	145813981	cg27647370	0.672785333		22	42686055
cg11517309	0.653408333	REEP1	2	86565390	cg27647559	0.657394		11	58420680
cg11521799	0.521395667	SNRPB	20	2451663	cg27648020	0.712784667	ATAD2	8	124408845
cg11532863	0.733687667				cg07289306		MIR138-1		
cg11546251	0.678851	ZNF552	19	58322078	cg08422803		ITGB2		
cg11572675	0.735643		2	86155952	cg14010194		GUCA1B		
cg11575912	0.742418333	NRTN	19	5828299	cg01656216		ZNF438		
cg11597157	0.654989333	KIF7	15	90171643	cg07553761		TRIM59		
cg11601336	0.664100333	HOMER3	19	19053012	cg09809672		EDARADD		
cg11615347	0.661998333	ENPP7	17	77710631	cg10281002		TBX5		
cg11631871	0.717051	UBTD2	5	171710573	cg15724772		ACOT7		
cg11652982	0.675385667		10	773101	cg11919692		Intergenic		

Appendix A

cg11653466	0.533732	C8orf86	8	38386235	cg08345526	ITPKB
cg11656992	0.732997667	P4HB	17	79818741	cg01552966	DNMT3A
cg11663289	0.681868333	PPM1L	3	160473298	cg06528771	Intergenic
cg11690666	0.686267667	NARF	17	80415469	cg12206840	Intergenic
cg11716578	0.681331667				cg10650214	LZTFL1
cg11725182	0.717616333	NPHP4	1	5968678	cg13373048	Intergenic
cg11738208	0.689229333	ZFPM2	8	106620533	cg03733470	SENP5
cg11744351	0.749650333	UBL3	13	30424153	cg04725636	DNAJC5B
cg11754670	0.659329333	TOR1AIP1	1	179852195	cg12066594	PSMB7
cg11762703	0.654089667	RPTOR	17	78929938	cg24501073	ABTB2
cg11784742	0.676494	SNCG	10	88722838	cg02042026	ZBTB16
cg11797364	0.695271333		6	436969	cg01381203	FOXM1;RHNO1
cg11817978	0.587855	C15orf39	15	75494295	cg02175213	CPM
cg11823253	0.655749667	VASH2	1	213154553	cg06084585	DLEU1
cg11828669	0.647391333	ZBTB1	14	64971476	cg07252680	SERPINA1
cg11832281	0.660070667	CUGBP2	10	11211022	cg09967176	Intergenic
cg11853000	0.709767	ZNF763	19	12075035	cg17261469	FBXO31
cg11858249	0.719266	HSF2BP	21	45079790	cg22508829	TOM1L2
cg11872478	0.709876667		2	96196596	cg22146772	Intergenic
cg11877270	0.747541667	SPRED2	2	65658583	cg13064873	TMEM92
cg11891330	0.552503333	WHSC2	4	2010714	cg13822123	PSME4
cg11908057	0.561217	HOXA4	7	27171154	cg02449373	FUT1
cg11955541	0.595814333	PDE4DIP	1	145040160	cg02683350	ADAMTS2
cg11967952	0.661349	TXNDC12	1	52498983	cg26042024	ZFAT
cg11973777	0.742479333		8	8244085	cg22871797	CYFIP1
cg12041266	0.662843667	ARRB1	11	75050147	cg18598861	IRF9
cg12045237	0.703638	TREML2	6	41169171	cg09777776	ZNF254
cg12054648	0.659214667	PPP2R1A	19	52692935	cg20545941	MPPED1
cg12058781	0.720711667	GBP4	1	89663896	cg19935845	TNXB
cg12062133	0.662930333		8	142548839	cg24423782	MIR182
cg12071328	0.720498	NELL1	11	20690930	cg19227382	CDH23
cg12078092	0.663507	ACOX3	4	8378947	cg03467256	HPCAL1
cg12100751	0.747304667	C1orf59	1	109203672	cg25196881	THBS1
cg12100956	0.677049333	GAA	17	78086420	cg02321112	MXN1-AS1
cg12121166	0.661656333		11	2376275	cg00355799	LOC339529
cg12133425	0.736506667	ZNF496	1	247494926	cg17556588	PRRG4
cg12137450	0.656317667				cg22618720	MIR5095
cg12160011	0.627164667	REV3L	6	111804095	cg24318598	ANO1
cg12170787	0.595774	SBNO2	19	1130965	cg07015775	ZNHIT6
cg12191232	0.675945	KIF19	17	72345492	cg21018156	LINC01312
cg12219045	0.742150333	LDOC1L	22	44894262	cg07475527	RCAN3
cg12250496	0.663828333		4	62066787	cg20000562	SFTA3
cg12253859	0.699849333	CYP2C9	10	96748928	cg07436807	STAMBP1; ACTA2
cg12275289	0.702389	TUBB	6	30691665	cg14029912	BHLHE40
cg12285326	0.669095667	APLP1	19	36365557	cg10812236	PLEK
cg12393668	0.672595333	CCDC43	17	42768503	cg22512011	KCTD12
cg12394201	0.659140333		11	43942418	cg24914185	FTCDNL1
cg12394289	0.700217	EHMT2	6	31856706	cg17651972	NXPH4
cg12396368	0.674961667	ICMT	1	6294850	cg11498967	C1orf198
cg12412575	0.675335667	AKAP13	15	86167192	cg10307212	TEC
cg12419482	0.669741333	P2RY2	11	72933282	cg06897548	TAB1
cg12425673	0.634369	WNK1	12	862975	cg13958199	NEK6
cg12426680	0.693262333		1	59282482	cg05710777	LINC02245
cg12510708	0.657501667	NFE2L3	7	26193805	cg02079181	RHEX
cg12516875	0.692940667		22	46463543	cg04872689	PLEK
cg12552944	0.704428333	SNX8	7	2353682	cg13980719	TNP1
cg12582330	0.680491667		1	2473224	cg02491017	HOXC4
cg12585429	0.727914333		16	49889252	cg26215428	PKD2
cg12586707	0.729587667		4	74738902	cg14216068	HOXA3
cg12614667	0.676779	GBP4	1	89663868	cg20163085	APITD1
cg12646649	0.654388667	LBX1	10	102987257	cg08732950	CBFA2T3
cg12662929	0.589776		1	160368823	cg07899076	RNF216
cg12673559	0.669808667	INPP5A	10	134590772	cg06639874	MLPH

Appendix A

cg12679910	0.530683333	MRPL38	17	73900713	cg15006881	GDF6
cg12686055	0.660283667	ANO6	12	45628262	cg05861567	MLC1
cg12728240	0.699698	NAT8	2	73870294	cg11573170	DIP2C
cg12731773	0.661010667	TSPAN4	11	846055	cg09591524	HOXA3
cg12766383	0.536616333	UBR4	1	19403306	cg16374343	ABR
cg12815987	0.723697667	GPR68	14	91719608	cg12935350	MRPS9
cg12852151	0.65583		6	136547799	cg05575921	AHRR
cg12856965	0.653929333	CROCC	1	17284740	cg23006040	LOC339975
cg12885832	0.553443	SH2B3	12	111843885	cg03725309	SARS1
cg12911791	0.730838667	CSHL1	17	61988811	cg14491535	JAZF1
cg12923994	0.669911667				cg05228408	MTHFR
cg12928619	0.582454333	WDR54	2	74648754	cg23933602	RSU1
cg12929983	0.668539667	BMF	15	40399004	cg21566642	ALPG
cg12933431	0.663007667	ATP10A	15	25959239	cg01620164	FIGN
cg12955084	0.664034	TMEM91	19	41886265	cg12479512	RBSN
cg13033858	0.657813667	SSH1	12	109248326	cg16604233	COL11A2
cg13072940	0.658817667	MON1A	3	49967521	cg22862003	MX1
cg13077699	0.705829333	ERLIN2	8	37593835	cg08860619	UBE2MP1
cg13086606	0.660328	MIR548H4	15	69452796	cg12303981	GMD5
cg13089335	0.675373	POU2F1	1	167203243	cg10455785	HDAC11
cg13091456	0.653247	TREH	11	118550379	cg15983626	GPC1
cg13119578	0.661291333		11	65196696	cg07964553	NEUROG2
cg131339542	0.597091333		2	8242815	cg22293458	VPS8
cg13175981	0.671737333	MCL1	1	150552382	cg12671121	STK19
cg13200739	0.655142	SHANK2	11	70805534	cg26955383	CALHM1
cg13256546	0.696729	APOBEC3G	22	39472544	cg22454769	FHL2
cg13267368	0.694689333	KLHDC2	14	50234727	cg03129964	BARX1
cg13294594	0.714969667	MAST1	19	12949443	cg06223172	CBLN1
cg13356370	0.659344333	CILP	15	65504725	cg01297357	ASIC1
cg13360866	0.679716		1	1135492	cg25950235	MIR9-3
cg13375457	0.668790333	RGL2	6	33263343	cg09926486	FRMD5
cg13385114	0.686239667	TMEM201	1	9656493	cg08622677	PRMT8
cg13390975	0.744900333	BRIX1	5	34915890	cg05190790	SOX1
cg13393782	0.669796	PRDM16	1	3099804	cg09476997	SLC9A3R2
cg13398782	0.606771	E2F2	1	23857288	cg12920180	COCH
cg13406003	0.744340667		6	127535477	cg13543355	LRRC74A
cg13418632	0.678973667	AP3B2	15	83332594	cg03312958	IRX2
cg13423282	0.748546667	SLC25A42	19	19174731	cg02003183	CDC42BPB
cg13433446	0.745535333		9	128170534	cg07600636	RASSF10
cg13446584	0.669332		7	74024953	cg19006008	F2RL3
cg13452386	0.662195333	NUDT8	11	67398435	cg08067617	F2RL3
cg13457700	0.657593667	ASB13	10	5708797	cg08200625	F2RL3
cg13468174	0.736113333	ZNF584	19	58919984	cg14021375	F2RL3
cg13513498	0.689687		8	102170922	cg15233062	HOXC4
cg13521124	0.681614333	TMEM151A	11	66064082	cg15700739	HOXC4
cg13559099	0.556931	TMEM179B	11	62554740	cg22198132	HOXC4
cg13575205	0.673090333	C7orf50	7	1039620	cg14108840	HOXC4
cg13583088	0.659874667		14	94360154	cg23697546	HOXC4
cg13602484	0.697691	MIR642	19	46177353	cg19186380	HOXC4
cg13604008	0.685115667	GTF3C5	9	135906154	cg05408649	HOXC4
cg13605327	0.733232		10	47057436	cg21493516	HOXC4
cg13606994	0.656173				cg27138204	HOXC4
cg13629563	0.716714	RUNX3	1	25256949	cg26201952	HOXC4
cg13645530	0.570002		12	116756948	cg22370252	HOXC4
cg13655563	0.675479333		17	1132368	cg18473521	HOXC4
cg13675015	0.682691667	TTBK1	6	43217925	cg03146625	HOXC4
cg13687834	0.65761		10	3514783	cg15648389	HOXC4
cg13696351	0.712392667	LRRC26	9	140063617	cg00243574	HOXC4
cg13698125	0.670404				cg01683044	HOXC4
cg13705391	0.656172	PRRT1	6	32118441	cg06714180	HOXC4
cg13716829	0.623060333	NRARP	9	140196785	cg10005224	HOXC4
cg13747275	0.702494667	BTBD2	19	1992952	cg18843682	HOXC4
					cg26035702	HOXC4

Appendix B

GENE	GENE	GENE	GENE	GENE
ETV6	CYB561	TPST2	C14orf72	LAPTM4B
C4orf44	TNFRSF8	MAP3K5	IL17RE	HGD
MGC23284	PTMA	TMBIM4	MTMR3	GBX1
NUDT17	ARRDC2	CHD3	MLLT1	JAK3
MFAP2	CORO1A	PCM1	MLPH	MOV10
MGRN1	KIF3C	CELSR1	SGK1	PVR
YY1AP1	STK40	PRDM1	CHST11	KCNE3
C1orf122	FADD	FBXL5	TCF19	TSGA14
RPTN	NAPA	SSU72	ANKRD11	UQCR
PGS1	TPM4	LOC100190939	KRTAP12-1	GSK3A
FBXO21	C3orf71	ZNF323	SLC7A11	FREQ
TMIE	GDPD3	TRIM3	MATN1	DLGAP2
STK10	ISCA1	OTUB1	ALKBH4	JMY
LRRC27	DBH	GPR44	IQSEC1	ERC2
C20orf43	CANX	TNRC18	SLC12A8	SCOC
ZNF527	LIMD2	ACTA2	ATM	PBX3
VPS28	NLRC5	SFMBT1	C17orf64	PER3
GOLGA7B	POLH	CR1L	C16orf87	JAKMIP1
SOHLH2	MCF2L	ADSS	TBC1D1	TNNT2
RECQL5	ABP1	LOC440926	ZP2	CDC73
ZNF365	FAM100B	RGS14	EIF4A1	C11orf2
CD300LG	KCNA3	RBM39	BRD1	KCNK12
ARHGAP10	FBXO7	CLU	PRR15	AFAP1L2
TRA2A	SIPA1L2	FGF1	CTNNA2	ZFPM1
ZNF423	SYNGR1	C20orf118	HMX3	KIAA0368
USP4	FAM86A	ARHGAP9	FAU	SCN4B
LTV1	F11R	C19orf60	HOXB2	ADCY5
ZNF518B	ERBB4	ELAVL1	JUB	POLR2I
ANXA6	MAP2K5	C9orf69	FAR1	SBNO2
ELOVL3	POM121	RNF216	CALHM3	TTC7B
RAB5C	TAF11	EGFL8	CREB3L1	RAPGEF6
RAB6B	LRRC8B	TBR1	CXCR1	FCHO1
ZNHIT6	ZNF121	E2F1	TSC2	GCH1
RBM19	MFSD2B	RHOV	KRT7	PPT2
RNH1	LOC338799	ALG1L2	KLF10	NUP153
TGFA	CNPY2	TMEM93	WRN	CYFIP2
ALS2CL	ANAPC11	ATAD3C	CENPA	ANGPTL2
ABLIM2	FRY	SERTAD3	SMARCA4	FKBP5
CPNE5	SLC2A11	TACC1	CORO2B	RAP2A
ICMT	UBE2F	ZCCHC10	LOC401052	COBRA1
RWDD1	C12orf66	LPPR2	ZBTB12	C9orf167
SPATS2	SYNPO	TMEM115	NDFIP1	PHGDH
PTGDS	ZBTB16	SLC27A3	C4orf29	ACAP3
SNORD23	SPARC	DAK	AFF3	TCF12
CPT1A	NPY5R	SPR	RECQL4	FGFBP2
IGFBP5	CDYL	MFSD10	ANP32A	PTPRD

Appendix B

DPRX	MKL1	MNT	CYP3A43	C5orf62
PCMT1	GPR176	NEUROG1	KIAA0892	NCEH1
DKK2	C4orf23	FBF1	FAM91A1	HDAC4
ZEB2	C6orf27	C10orf78	FCGBP	GMPS
RPAIN	CCDC61	UNC84B	PINK1	RABEP1
ZFH3	FKBP9	HERC2	RNF34	MAMDC2
EBF3	MAP7D1	SETBP1	KCND3	LARS2
MGMT	SLITRK5	DPY19L2P4	RNF25	SIK3
KIAA1543	IL15RA	MOBK12A	ARHGEF10	CXXC5
ZNF48	ABCF2	PXN	TTYH3	PRPSAP1
BAHCC1	ABCF3	PDE2A	KIF1A	SELO
ZFP161	USP13	PSMB9	PARD3B	ST14
TCP11L2	SRRM2	C1orf130	ZNF574	TMEM61
PDE10A	MICALCL	PLEKHA6	YAP1	MYEF2
RGS6	HK1	LLGL2	RDH10	ITGAE
BDP1	VAR52	ALKBH5	TBC1D9B	EDN2
PCIF1	LPXN	FAM18B2	SLC35C1	C1orf200
RET	TNFAIP3	LOC619207	PIGY	C2orf60
LOC728640	ST5	GPS2	PTPRN2	SLC12A9
CORO1C	AEBP1	RASD1	CUEDC1	C21orf33
LOC728743	BTRC	EIF3M	SFRP5	XRCC1
SLC7A5	ANK1	LRRC42	SHBG	ZNF714
NANOS3	PTPRE	HEY2	NTN1	TPRX1
CLK2P	VAC14	DIP2C	PLCH2	FOXK1
MIR802	RBP7	AHNAK	CREG1	COL9A3
MIRLET7I	SNTB1	PDP2	DGKZ	CAMK1
VWA3B	RCHY1	HIVEP1	MAP7	F12
MAN1C1	SCP2	VP52	SH3RF3	RAVER2
SNRNP25	GJA9	GPR177	DLL4	DDX10
PRICKLE4	NDUFV1	ARHGEF2	SFRP1	TMEM82
C1orf9	RNF26	MSH5	GLIPR1L2	ADAM11
RCC2	SOX8	ATG16L2	SAP130	CASP2
DAZAP1	HSDL1	TNFRSF13B	FASTKD5	C17orf51
FAIM	C7orf50	MEIS1	C16orf59	RTKN
BICD2	CBFB	OXER1	MACROD1	EP400
MARK2	FAF2	MIR29C	GP1BB	GNB1L
REEP3	SIRT5	PEX10	NISCH	ASB10
C10orf4	JARID2	SOX1	NTRK1	STARD3
IL4R	ORC6L	SLC6A20	SERPINF1	COG1
HIF1A	RALGAPA2	VGLL2	LETMD1	MYO1C
ARID3A	MUM1	C12orf53	MSH3	CYC1
LOC150381	TMEM39B	MUC5B	XYLT1	WBSCR16
CEBPE	OXTR	ASCL2	SNORD119	VTI1B
DCK	CHCHD6	KATNAL2	CDT1	LPIN3
ALAD	ZP1	ZNF700	CSF1	RBM20
RRAD	SKI	FND5	OSBP	GDNF
KRBA1	RBM16	VENTXP7	NDUFA6	ZNF542

Appendix B

CCHCR1	ATP2B2	SYDE2	ACSL1	CACNA2D2
EVC2	HYI	TSPAN18	ZNF598	MAST3
VKORC1	HAL	MIR636	RNF126	FEM1A
SLC1A5	CD47	LTBP4	ABHD8	TFAP2E
HES1	PDE6A	PILRB	SLC25A19	NCRNA00171
C9orf93	KISS1R	LGR6	FLJ40504	BCL3
MAFA	HCCA2	ATF3	ZNF195	RAET1K
PPIL2	CASZ1	CD58	PSMD8	CYP1A1
RAB3B	P2RX3	MATN3	C22orf32	ESRP1
RNF220	KPNA2	C9orf122	PHF21B	SLC7A6OS
PLEKHG5	PLEKHF1	MIR769	KIF13B	RGPD2
SFRS6	SORBS1	BAI1	MMAB	DNAJC27
BANP	CPNE6	RNF208	C3orf63	ARHGAP20
PRX	ZNF692	COL11A2	CCND2	FHOD1
PRKCD	PLEKHO2	CCDC33	DGAT1	KCNQ1OT1
UCK1	MSLN	HNRNPR	IGDCC3	MICA
FOXK2	ETS2	PHOSPHO1	TAX1BP1	UXS1
ZNF84	LRRC43	CCL3	CDK11B	GLB1L2
HGS	POMC	CUX1	SYNGR2	RPS15
ETV5	RBPM5	ACSM4	SRD5A3	PWWP2A
PHOX2A	LOC440356	HLA-E	DLK1	LPAR5
SCLY	PCGF3	CSNK1D	C10orf137	IL17RA
FLYWCH2	FAM181B	MRPL42	LRRFIP1	LHB
D2HGDH	SLC44A4	LYST	SLC22A18AS	C7orf44
PITPNM2	GPT2	CAPN8	PRIC285	C10orf25
ABCC1	PTEN	ARID1A	IGSF22	TCF7L2
MACF1	GIPC1	ELAC1	TUBGCP3	SYT2
CDC20	DHCR24	NFKBIL1	FITM2	TNXB
C6orf136	SLC9A3R2	CPXM1	ME5DC1	UBE2L6
CHFR	ANKDD1A	KIAA1191	ATXN7L2	MYL6B
TMEM190	PLXNC1	LRP8	UBE3A	STAT5A
SENP3	GALC	TNFRSF4	HOXB7	FABP3
HAR1B	MCM7	DVWA	ZNF180	NOM1
KANK2	ARHGEF7	POLR3E	ZDBF2	TOMM22
TMEM184A	APRT	SLFN12L	BNIP3	KAT2B
C19orf12	EHBP1L1	SNX15	TMEM51	SETD7
NFKBIZ	LOC399815	CDKN1C	C16orf57	HECA
PLEKHA2	CYB5R3	CD68	STK19	BAD
CCT6A	IGF2BP1	HDAC11	CRTC3	UHRF1BP1L
ADNP	NAA25	SMCR7L	CDKL3	COG7
ECEL1	TRIM47	CKB	RHPN1	MAPK4
OAZ1	PTPN5	TLE2	AGAP1	IRGM
ATP11A	PPP4R4	CHD6	PRRT1	MORN1
ECE2	CDC25A	SLC30A2	PLEKHG3	CRLF3
TMEM131	ZNF790	SNAPC5	NME3	PLA2G4F
RPS6KA1	LRP5	NMNAT2	ZNF295	SLC10A7
AMDHD2	SLC12A2	SNORD46	MTUS1	GAA

Appendix B

PUF60	PCDHA6	C7orf25	SLC29A4	TMEM59
HOXA3	DTYMK	LMNB2	MYO1F	GGT1
STAB1	SCARB1	CD55	RASA2	PRKCB
CDC42BPB	ITGA3	TBC1D2B	NCK1	MALL
ZNF672	PNKP	JDP2	SORCS2	CLEC4D
HIP1R	HMGCR	AMPD2	CCNL2	PPP2R5E
PILRA	ZNF624	RTTN	CBLN1	TPO
SLC25A10	ZNF808	MST1	ADCK1	C1QC
SLC1A4	BCAS3	PAPSS1	CASP7	DHRS13
ACVR1	CLDN18	ZFR2	ZNHIT1	CLTB
FAM59B	H3F3A	MMP11	ZNF529	BAMBI
SNX21	AIM2	LARP1	SLC27A1	NBR1
MPO	CD164	TMEM90B	DCI	RAB11FIP5
OBFC2B	AP2S1	BRF1	KIN	LASS6
AGA	FGGY	CREB3L2	ZBTB46	CDKN1A
LOC100134229	CSTF3	MYO7B	C16orf13	MIR612
S1PR5	SPTBN4	RASSF3	SFMBT2	ERN1
VAR5	ARID1B	FLJ43663	PHB2	RNF4
LPAR3	EDEM2	HELB	PUSL1	PPARA
CLEC3B	PTDSS2	ACOT11	COLEC12	RPL10A
ATAD5	HOPX	TTLL1	WDTC1	ZBTB22
FBXL14	HCN3	MRPL12	AZI1	PRDM16
LOC441601	CYTH3	NINJ2	KIAA0100	ZNF764
HRCT1	ARL3	FAM125A	GGT5	PLBD2
RHOT2	INPP5A	ANUBL1	ZC3H3	RUNX1
H6PD	SREBF1	IL3	BHLHE23	ARL8A
ZMIZ1	TXNDC16	TMEM175	P2RX5	HDAC7
ZWINT	SF3B2	C10orf47	TUB	MIAT
RAB3D	PMF1	RASGRF1	PFDN2	MGC26597
ZNHIT2	C21orf15	TNFRSF10A	GABRB1	VPS18
QRICH1	CSRP1	FNDC3A	DLEU2	STAT3
ZNF592	NANP	B3GALT4	PAIP2B	ITPKB
ONECUT2	LTBP3	KCNN4	CMTM3	TMUB2
SLC35E1	MAN2A2	CENPM	NHLRC4	ABHD14B
HSD17B12	LTBP1	TBCA	RPIA	CLIC6
MAD1L1	ADD2	MIR145	EIF2B4	ERH
MYH9	MCAM	MMP2	LOC441046	TMEM151A
IL21R	GLYR1	IREB2	DYNLRB2	TMEM179B
ATE1	FOLR2	SMAD5	PPP4C	MIR642
FUT11	PRDX5	ZBTB9	SV2C	GTF3C5
ARMC8	CCNI2	DNAJC11	ZNF541	RUNX3
TTBK1	EPHB3	MYOM3	DEPDC7	LRRC26
ACBD3	SYT8	PRR7	C20orf117	NRARP
RPS21	UGT8	LOC642006	TBC1D14	BTBD2
SLC16A14	LOC729991-MEF2B	PYCR2	ABCG1	RAPGEF1
AMFR	BLM	KRAS	FOXJ3	VIPR2
SEMA6B	RLN3	MUC6	IGFBP6	ZCCHC24

Appendix B

TTF1	B9D1	MIR7-2	CASR	TGFBR1
WDR81	NPC1L1	TSSC1	PANK2	MUTYH
HMHA1	KIAA1688	TPCN2	IGF1R	ST6GALNAC4
SELK	REEP1	CHRNA3	LOC91450	ABL1
TRNAU1AP	SNRPB	TMEM132D	AKAP13	NOTCH1
DGCR9	ZNF552	ODZ2	CNOT8	NKX2-4
CEP192	NRTN	NCL	KCNQ1	ZNF174
LMF1	KIF7	CUL9	PNPO	ATL2
PRPF6	HOMER3	PUM1	ARHGEF16	ZNF143
PSMB11	ENPP7	EPB41L1	ITIH5	CHD7
THRA	UBTD2	KCNK7	PHACTR3	LOC100270746
CRISPLD1	C8orf86	FOXO1	MAEA	C20orf114
C14orf43	P4HB	MYO15A	AHRR	HIP1
EPHX3	PPM1L	EDEM1	PPP2R2B	CLK1
ECHDC2	NARF	RACGAP1	TXNL4B	ABCA2
C16orf80	NPHP4	UNCX	KLHL30	APOE
PGM1	ZFPM2	KCNN1	PRDM11	SECISBP2
SOCS1	UBL3	LOC647979	HSPA8	SGK3
MIR548F5	TOR1AIP1	SLC30A3	MAPT	MAPKAPK2
MRPS7	RPTOR	KIAA0195	JMJD8	SSB
ESD	SNCG	KRTAP5-7	STRADA	MGC45800
MIR589	C15orf39	MAPK1	TRAPPC9	NFIX
FAM8A1	VASH2	LEPR	SFRP2	CTSL1
GARS	ZBTB1	PTH1R	TAAR6	RIN1
OAT	CUGBP2	WAPAL	LRFN3	BMP8A
IQCE	ZNF763	RNASEH2B	MRM1	CHAC1
ALPP	HSF2BP	GALNT5	PROX2	AKAP1
AP2B1	SPRED2	RASSF4	NTN3	PDP1
HS6ST1	WHSC2	KDM2B	KIAA0232	HMGA1
DBP	HOXA4	C1orf177	ECHDC3	FAM184B
TRIM15	PDE4DiP	RAB21	ACACB	NRP2
SPTAN1	TXNDC12	PIM3	EPPK1	RNASEH2C
VAPA	ARRB1	WWTR1	KHDC1	HMGXB4
RTN4RL2	TREML2	MARK3	ZBTB4	TBCEL
ZIC5	PPP2R1A	RHOF	SNORA38	AMOTL1
KIAA0226	GBP4	SLC18A3	BAG2	CMPK2
AP3M2	NELL1	TAP1	LPPR3	TNIP2
TYROBP	ACOX3	DLG5	PROC	DNM2
CIAO1	C1orf59	PLEC1	ARHGEF17	SLC30A5
CHMP2B	ZNF496	DOCK9	KIF26B	SLC03A1
FGFRL1	REV3L	ZNF787	SIPA1L3	PLAGL1
FAM76A	KIF19	RADIL	HEXIM1	MSI1
C1orf129	LOC1L	SOX18	BAT3	SOX12
IPO9	CYP2C9	PTBP1	ACTN1	BRP44
SMOC2	TUBB	DSCAML1	SSTR1	DDR1
ANKS3	APLP1	CYB5A	MTRR	C2CD2L
ZNF321	CCDC43	NOX4	PCDH24	ZBTB44

Appendix B

KIF1B	EHMT2	SLC6A6	VOPP1	PTF1A
SLC17A5	P2RY2	AKR1A1	MIR196B	WDR46
DEPDC6	WNK1	RXRA	ERGIC1	DNAJB6
GADD45GIP1	NFE2L3	CXCR5	ITGA9	RNF165
GLI2	SNX8	KDM5B	SPAST	PTCD2
ROBO4	LBX1	S100A6	P2RX7	ZNF589
EDNRA	MRPL38	MRFAP1	PIAS3	AIRE
SIGLEC10	ANO6	EFNA1	PPBPL1	CNNM1
SLC5A6	NAT8	DND1	TBL1XR1	CCDC72
CNR2	TSPAN4	FMNL1	MEIS3P1	CHSY1
C13orf37	UBR4	GFRA2	MEA1	C20orf3
F2RL3	GPR68	AGPAT9	BHLHE41	XRN2
NUDT1	CROCC	NLRC4	PPIF	SPIRE2
MICALL2	SH2B3	RAG1AP1	LOC400931	APOLD1
PSMG1	CSHL1	TBX4	SLC2A10	DOPEY2
LFNG	WDR54	SERPINA1	CCDC42B	UNC13A
EFNA5	BMF	DUS2L	SOX21	CYTH1
LZTR1	ATP10A	POLR3GL	ZNF320	PIK3CD
SARS	TMEM91	ZNF142	GPX4	CSNK1G2
DEGS1	SSH1	OSM	PCDHA2	PDIK1L
C1orf95	MON1A	TRPV1	ZC3H6	ALOX12B
KIAA1875	ERLIN2	CORO7	PXMP2	SLC9A3
KCNK4	MIR548H4	STK11	CS	ATAD2
SNED1	POU2F1	KLC2	BOC	C10orf125
KRT81	TREH	NR2C2AP	SDHA	TAP2
MTFMT	MCL1	PRKCZ	S100Z	RNF130
DDC	SHANK2	CHRM2	L3MBTL4	CNBP
PTCH1	APOBEC3G	CARS2	SNUPN	GUCY2D
CCDC18	KLHDC2	NFIC	COX4NB	RRP9
AMACR	MAST1	CHRN2	PAK1	ARF1
BAIAP2	CILP	FAM22F	C17orf95	SOCS3
KLRC1	RGL2	ZNF43	SP4	GFAP
SLC38A10	TMEM201	CYP51A1	REXO1L2P	NAGLU
DCTPP1	BRIX1	ANKRD33B	ADRB2	ETHE1
ZNF688	E2F2	GTF2I	TTC38	SUPV3L1
PLXNA2	AP3B2	TRIM50	IFT122	MOGS
USP34	SLC25A42	ACCN1	SDK1	ZNF224
TCEB2	NUDT8	CFLAR	WIZ	ZNF652
ANKRD29	ASB13	LBH	PRKRIR	IL18RAP
RHEB	ZNF584	RANGAP1	NR2F6	SGTB
DUSP27	PTPN1	ALOX5AP	EPHA8	NDN
EMB	DHX34			

Appendix C

ILMNID	mean_cv_auc	ILMNID	mean_cv_auc	ILMNID	mean_cv_auc
rs10011838	0.519726333	rs9289146	0.526316667	rs7687819	0.520932333
rs1003081	0.517051667	rs9293518	0.537482333	rs7699288	0.518790333
rs10057967	0.514961333	rs9293683	0.526262667	rs7703051	0.516950333
rs10061288	0.510524333	rs9294148	0.523948333	rs7704555	0.518022333
rs1006195	0.518588667	rs9298076	0.514307333	rs7714712	0.512168333
rs10077431	0.519279	rs9298506	0.520318	rs7721099	0.513771333
rs10077875	0.514765667	rs9303277	0.511603333	rs772178	0.520062667
rs10086783	0.519548667	rs9309245	0.532550667	rs7780752	0.520198
rs1008723	0.516102	rs9310427	0.515816667	rs7786095	0.510560333
rs1009017	0.527636	rs9312517	0.511719667	rs780092	0.515400667
rs10091344	0.514352333	rs931619	0.541074	rs7835508	0.515336667
rs1009188	0.524421333	rs9317220	0.514578333	rs7845800	0.514660667
rs10093110	0.536918667	rs9321870	0.531881333	rs7860634	0.523839667
rs10100760	0.515994333	rs9327000	0.522663667	rs7865618	0.523330667
rs10107182	0.519885333	rs9327455	0.511283	rs7870970	0.516371333
rs1011731	0.512390667	rs9329223	0.512433667	rs7874142	0.530665333
rs1012089	0.552717	rs9332408	0.520188333	rs7899106	0.510139667
rs10125521	0.517637333	rs933769	0.546878	rs7901695	0.511570667
rs10143198	0.530187667	rs9346633	0.517316333	rs7901883	0.519445667
rs1015291	0.52184	rs934778	0.525646	rs7912454	0.522888667
rs10158937	0.518234667	rs9349211	0.511000667	rs7922045	0.521223333
rs10160382	0.514322333	rs9355878	0.518065333	rs7928810	0.515534333
rs10172070	0.533763	rs9357402	0.520871	rs7928968	0.547947667
rs10180663	0.510354	rs9357506	0.517623333	rs7931302	0.514699667
rs10184839	0.513977333	rs9358928	0.514004333	rs7932891	0.514611333
rs10189761	0.519119333	rs9368716	0.511324333	rs7943936	0.516721
rs10192654	0.524167	rs9375695	0.526935667	rs7963783	0.517982
rs1020410	0.542570333	rs9376353	0.520625	rs7968902	0.512004667
rs10204803	0.515904333	rs9379130	0.528216667	rs7987314	0.520261
rs1020731	0.535229667	rs9384878	0.537997333	rs7993238	0.522138
rs10208649	0.523956667	rs939479	0.510348	rs7998314	0.533312333
rs10233389	0.539814333	rs939583	0.515296667	rs8017825	0.519430667
rs10253361	0.513503	rs9408882	0.522792333	rs8032301	0.542065
rs10256972	0.516863	rs941853	0.516868	rs8033957	0.516029667
rs10279261	0.511418667	rs943346	0.514644333	rs8042007	0.514718
rs10282458	0.522328333	rs943466	0.533034	rs8043389	0.528941667
rs1037925	0.518734	rs9435732	0.521202333	rs8049033	0.560152667
rs1042725	0.518518	rs9435739	0.529193	rs8051363	0.535149333
rs1044531	0.520767	rs943764	0.527382667	rs886750	0.519450333
rs1044826	0.510021333	rs949540	0.511049333	rs887829	0.516612667
rs1045530	0.523788333	rs9505118	0.521612333	rs888192	0.526085667
rs10459592	0.546661667	rs950776	0.512702	rs891124	0.520673
rs10462509	0.510782333	rs950802	0.523879333	rs891903	0.514889
rs10482795	0.514941667	rs9508025	0.536558	rs9009	0.518450333
rs10489318	0.538336333	rs951266	0.522497667	rs907089	0.520681333
rs10491208	0.514447	rs951366	0.517672333	rs907183	0.537053333

Appendix C

rs10491334	0.519718333	rs9515905	0.523432667	rs920590	0.521254
rs10493442	0.519077	rs9517320	0.512585	rs9267954	0.512453667
rs10494964	0.510181333	rs9538141	0.516687333	rs9268402	0.524000333
rs10497251	0.510063333	rs9540493	0.511198667	rs9273363	0.519030667
rs10498051	0.513567333	rs954768	0.514032	rs9284956	0.516694
rs10498767	0.517472667	rs9548050	0.521162	rs9285458	0.517726
rs10498891	0.510963333	rs9600236	0.524330667	rs928579	0.529841667
rs10501174	0.516428333	rs9620115	0.521978333	rs9287993	0.520800333
rs10503186	0.513158667	rs9634445	0.521822333	rs8055190	0.510600333
rs10503669	0.524496333	rs9645860	0.522857667	rs8055236	0.541007667
rs10505725	0.547836333	rs9660067	0.514454667	rs8068318	0.514241
rs10506971	0.534997333	rs9678859	0.511386	rs806851	0.528952667
rs10508802	0.520966	rs974266	0.512751667	rs8071463	0.524138
rs10510110	0.528298333	rs978332	0.523032	rs8079727	0.517233333
rs10510419	0.511338333	rs9816029	0.514303333	rs8084351	0.524236333
rs10511377	0.553506333	rs9818870	0.533084333	rs8093407	0.524506667
rs10511378	0.52949	rs982077	0.510978667	rs8096662	0.521477333
rs10513688	0.530724333	rs9838625	0.515808333	rs8102873	0.516038667
rs10514302	0.510754333	rs9840967	0.520438	rs8141797	0.511209333
rs10517653	0.521984	rs9845966	0.53483	rs8142788	0.516341667
rs10517660	0.517841667	rs985060	0.511159667	rs814953	0.555297667
rs1053711	0.511885	rs9856633	0.518132	rs8181588	0.522044667
rs1057558	0.511700333	rs9859077	0.512242	rs828544	0.531733667
rs1060435	0.514240333	rs986363	0.512862	rs833520	0.522940667
rs10732812	0.521043667	rs9890032	0.532820333	rs836525	0.521462667
rs10740094	0.526927	rs9895032	0.513745667	rs844767	0.525469333
rs10742752	0.525057333	rs9920066	0.528427	rs849141	0.510455
rs10745460	0.516834	rs992072	0.528285	rs852787	0.519617333
rs10746862	0.516815	rs9927842	0.513908	rs862049	0.543632333
rs10753801	0.543639333	rs9940128	0.511276667	rs739401	0.512912
rs10755578	0.517295333	rs994270	0.517485	rs7395662	0.551445667
rs10757272	0.533145	rs9944249	0.523139667	rs7396835	0.519438667
rs10757283	0.520753	rs9951893	0.526531	rs7396851	0.521221333
rs10760361	0.527261333	rs10004128	0.523359333	rs7400722	0.529214667
rs10765792	0.525382333	rs10012103	0.534649	rs741486	0.540244667
rs10769908	0.521146333	rs1001789	0.525431333	rs741959	0.536305667
rs1077181	0.5167	rs10044122	0.528074667	rs743236	0.513821
rs10777237	0.525459333	rs10054402	0.52609	rs743757	0.529727667
rs10777845	0.533639667	rs10057809	0.540879667	rs745570	0.518362333
rs10786679	0.516016667	rs10069629	0.539153333	rs746463	0.522315667
rs10787287	0.517984667	rs10071895	0.531081333	rs746630	0.521117667
rs10789340	0.512370333	rs10076833	0.519814667	rs7481311	0.539417667
rs10789752	0.523041	rs1007775	0.531469333	rs7513580	0.543248667
rs10790519	0.523512667	rs10102276	0.529349	rs7537765	0.516017667
rs10790800	0.515501	rs1010376	0.543323333	rs754133	0.542253667
rs10793514	0.514592	rs10138608	0.543924	rs7544372	0.512256333
rs10804990	0.523516333	rs10162686	0.519836333	rs7544735	0.518262

Appendix C

rs10821808	0.536531333	rs10166894	0.545665	rs7549339	0.519192667
rs10823559	0.515362333	rs10212337	0.529956333	rs756717	0.515968
rs10824026	0.513683	rs10215031	0.521788667	rs7571469	0.549634333
rs10824347	0.517859	rs1022361	0.501795333	rs7572970	0.532363333
rs10824378	0.511435667	rs10252758	0.509564	rs7573672	0.511710667
rs10829156	0.510189	rs10270572	0.527245	rs7575189	0.514084333
rs10830962	0.536053667	rs10275475	0.534347333	rs7581601	0.510276333
rs10832255	0.510581333	rs10275897	0.521574	rs7583236	0.520472333
rs10838465	0.5329	rs1028990	0.532971667	rs759250	0.517848667
rs10838681	0.517722	rs1034851	0.546055333	rs7593730	0.526388333
rs10840060	0.518881667	rs10402812	0.525249333	rs7600417	0.515158333
rs10840100	0.510369667	rs10420793	0.540761333	rs7609422	0.511888333
rs10842702	0.539893333	rs10462018	0.530736333	rs7622114	0.519891333
rs10846606	0.516683333	rs10483088	0.516385667	rs7637852	0.510540333
rs10850071	0.524228667	rs1048546	0.532053667	rs7649045	0.524038667
rs10851458	0.534101667	rs10485975	0.529680667	rs765547	0.511111667
rs10871590	0.534901	rs1048687	0.541733333	rs7669833	0.516039
rs10873298	0.518314667	rs10490778	0.52352	rs7672622	0.513528333
rs10886017	0.522871333	rs10493634	0.545307333	rs9537312	0.540654
rs10886863	0.510449667	rs10493947	0.532054667	rs9538523	0.524669
rs10894670	0.536822	rs10494190	0.542281667	rs9543704	0.520089
rs10900325	0.518824	rs10498209	0.526237	rs9553516	0.520333667
rs10903129	0.523496667	rs10500802	0.521349333	rs9556795	0.527008
rs10904908	0.520383333	rs10501898	0.52634	rs9578667	0.525602
rs10906115	0.516339667	rs10502353	0.536716333	rs9597353	0.528237
rs10915840	0.524777333	rs10503070	0.533309333	rs9697131	0.532691667
rs10919387	0.541232333	rs10505174	0.531548333	rs973309	0.531916333
rs10923654	0.510011	rs10505508	0.555358333	rs973406	0.54787
rs10923931	0.512017667	rs10506806	0.532483	rs9771401	0.532829
rs10928240	0.533504	rs10508223	0.531656667	rs9808028	0.525461333
rs10932444	0.52816	rs10509116	0.526403333	rs9817739	0.542757
rs10943567	0.523255	rs10511128	0.539279	rs9822586	0.539233667
rs10951191	0.513439	rs10517232	0.539193	rs9829295	0.526252667
rs10954284	0.514206667	rs10518436	0.52679	rs9895790	0.524472667
rs10957985	0.533645667	rs10518844	0.531969333	rs991035	0.519965333
rs10959841	0.521022	rs1062033	0.538765333	rs991088	0.541034
rs10965245	0.520976667	rs10743413	0.546806333	rs9915191	0.521000667
rs10965780	0.516968	rs10746129	0.530341667	rs9927884	0.526547667
rs10972619	0.522657333	rs10755048	0.532812667	rs992890	0.524578
rs10981835	0.533009667	rs10764566	0.534403333	rs9932117	0.524586333
rs10984755	0.532223333	rs10764712	0.519289333	rs9933953	0.522566333
rs10984756	0.532999333	rs10765092	0.523845667	rs9942926	0.520281
rs10992471	0.523034333	rs10774819	0.529042333	rs9957245	0.527680667
rs11001963	0.527349333	rs10777160	0.526515	rs9977055	0.517690333
rs11024074	0.517012333	rs10777556	0.528058333	rs9979091	0.533450333
rs11030102	0.528155333	rs10779613	0.527838333	rs9988960	0.554126333
rs11030104	0.515440667	rs10791538	0.534252	rs9993371	0.534667

Appendix C

rs11030119	0.526490333	rs10793169	0.527778667	rs864745	0.534474667
rs11031791	0.512378	rs10795997	0.489353667	rs868155	0.510354
rs11037685	0.536820333	rs10797056	0.496269333	rs869402	0.514078
rs11041699	0.518466667	rs10807242	0.518671	rs876383	0.522055333
rs11044977	0.517831667	rs10807948	0.547413	rs881301	0.539171667
rs11050349	0.532320333	rs10810175	0.542393	rs884847	0.522593333
rs11057072	0.515968333	rs10811875	0.524364667	rs886126	0.515776667
rs11057401	0.527379	rs10817270	0.533766333	rs7127355	0.512132667
rs11057408	0.524401	rs10828879	0.530463667	rs7132055	0.520522
rs1105956	0.522636333	rs10845724	0.518369667	rs7138803	0.528093667
rs11063069	0.523924	rs10847204	0.537559333	rs7140766	0.524029333
rs1106518	0.530971333	rs10847647	0.528126	rs7144011	0.530443333
rs11065898	0.512699	rs10847957	0.524924333	rs7168365	0.551081333
rs1106683	0.524029333	rs10848829	0.531640667	rs717662	0.518689333
rs1106684	0.531780667	rs10866272	0.527882	rs718545	0.533096333
rs11067763	0.513494	rs10869631	0.531039667	rs7188054	0.51257
rs11073147	0.531115333	rs10869657	0.533383	rs719727	0.519279
rs11074446	0.532383333	rs10873910	0.528129333	rs719728	0.511619
rs11082385	0.516359333	rs10877771	0.532017333	rs7213273	0.514574
rs11082402	0.517992	rs10894590	0.531362	rs7221773	0.513115
rs1109559	0.526890667	rs10898026	0.518115	rs7225461	0.529247667
rs11097470	0.524630333	rs10900579	0.533348667	rs7236548	0.510872333
rs11103390	0.529406333	rs10905492	0.517469	rs725519	0.522918667
rs11107116	0.512602333	rs10912994	0.529044	rs7255223	0.531145667
rs11123406	0.518746	rs10913672	0.544736333	rs725959	0.526316333
rs11124981	0.531283667	rs10916248	0.53508	rs7284873	0.517669
rs11127485	0.516279333	rs10924245	0.521769667	rs7285459	0.525566667
rs11131794	0.534314	rs10931023	0.529957667	rs7298565	0.523472
rs11135694	0.537263667	rs10939630	0.517105667	rs729949	0.516372333
rs11137826	0.523795	rs10948303	0.527385667	rs7299936	0.510814667
rs11144688	0.522502	rs10950957	0.525209667	rs730180	0.512683667
rs11153594	0.534755	rs10954327	0.532083667	rs7307780	0.510097667
rs11159091	0.520810333	rs10954681	0.544045	rs7307902	0.519258333
rs11159104	0.523715333	rs10954899	0.49402	rs7311105	0.511853333
rs11165468	0.523813	rs10957169	0.526446333	rs7313556	0.524225
rs11166986	0.525966667	rs10965113	0.522889	rs731681	0.515561
rs11169170	0.560848333	rs10966328	0.520239667	rs732679	0.525318333
rs11171739	0.523240667	rs10969027	0.496926667	rs7335249	0.521525
rs11175195	0.520504333	rs10976831	0.536876667	rs7360000	0.528912
rs11187837	0.530352333	rs10979145	0.522049333	rs9275772	0.543849667
rs11191548	0.51182	rs10992512	0.532370667	rs9275793	0.542875
rs11198898	0.510355	rs1099451	0.517049	rs9285256	0.537983
rs11204070	0.533573333	rs10996973	0.527936	rs9293713	0.522864667
rs11204791	0.516941667	rs10999736	0.528571333	rs9295990	0.527848333
rs1121980	0.511339667	rs11042458	0.545402	rs930167	0.525263
rs1122608	0.525696	rs11051643	0.548626	rs9304214	0.517130667
rs11229555	0.52833	rs11055616	0.532770667	rs9306113	0.514277333

Appendix C

rs1123320	0.511867333	rs11063455	0.527377667	rs9315638	0.532801
rs11242664	0.53037	rs1106789	0.536007333	rs9316670	0.534204333
rs11246136	0.512731	rs11072793	0.536427	rs9324789	0.548494
rs11247168	0.523792667	rs11074347	0.531670333	rs9328413	0.540944
rs11249913	0.525432333	rs11078208	0.535812333	rs9343171	0.547749333
rs11255908	0.513650333	rs11082523	0.526885333	rs9345148	0.526754
rs11258422	0.520061667	rs11090409	0.521912333	rs9345943	0.53492
rs11259474	0.533802667	rs1109670	0.530383667	rs9355674	0.533148333
rs1126089	0.511819	rs11098403	0.517893	rs9366337	0.563259667
rs1129555	0.510527667	rs11108495	0.53932	rs9393366	0.537221667
rs1153188	0.520684333	rs11109094	0.510932	rs9394789	0.543492
rs11570094	0.511739333	rs11110677	0.528073	rs9397922	0.536141333
rs1157755	0.523695333	rs1111850	0.542628667	rs9398716	0.533686667
rs11579440	0.51869	rs11120211	0.520989	rs9405214	0.538723
rs11591741	0.513118667	rs11125718	0.524453667	rs9406226	0.533407333
rs11652097	0.515416333	rs11156479	0.533453	rs9421389	0.531855667
rs11652437	0.510628667	rs11163296	0.526400667	rs9425569	0.535082667
rs11657101	0.514236333	rs11166722	0.522992667	rs9442385	0.525324
rs11662297	0.519545333	rs11166916	0.546963	rs9466768	0.521233
rs11662738	0.523553333	rs11185082	0.535517333	rs949280	0.527912
rs1169081	0.515837333	rs11196406	0.529596333	rs949944	0.528693333
rs11708540	0.510004667	rs11217583	0.526697667	rs9508344	0.529069333
rs11709077	0.518614667	rs11217584	0.526656333	rs9521469	0.529165667
rs11714208	0.524208333	rs11218825	0.526820333	rs9522321	0.527954667
rs11716980	0.524746667	rs11223991	0.52377	rs689183	0.520693333
rs11724872	0.515503667	rs11229889	0.519361	rs6893807	0.516628
rs11747001	0.522338	rs11237703	0.527716333	rs6899277	0.521686667
rs11760498	0.518766667	rs11239542	0.538384667	rs6899540	0.520520333
rs11767893	0.529912	rs11249990	0.545065333	rs6905288	0.513031667
rs11802413	0.523128667	rs11252717	0.538885	rs6907943	0.522438667
rs11830241	0.525082	rs1132200	0.535901667	rs6912853	0.53714
rs11897119	0.519724333	rs1133611	0.521260667	rs6920691	0.542576667
rs11899888	0.521207333	rs1149389	0.518405333	rs6921533	0.525662333
rs11929640	0.519914333	rs1154862	0.530279667	rs6922269	0.520638
rs11970772	0.515747333	rs1155204	0.531942667	rs6924570	0.526365333
rs11979110	0.513526333	rs1155741	0.529448667	rs6956675	0.517809667
rs1198430	0.520553333	rs11564026	0.526662333	rs6964415	0.511751333
rs1199338	0.528456667	rs11568820	0.523934667	rs6965195	0.516684
rs11997161	0.528695	rs11578006	0.5276	rs6965740	0.523417333
rs12044963	0.540881667	rs11591768	0.537764	rs6973609	0.527446
rs12047439	0.521698333	rs11592114	0.535555667	rs6983129	0.526704
rs12048743	0.516219333	rs11596249	0.52177	rs7008914	0.514895667
rs12049202	0.520494667	rs1159930	0.530131667	rs701815	0.518041
rs12052058	0.527632333	rs11602700	0.524594333	rs7024334	0.515094
rs12052201	0.527767333	rs11611897	0.533215	rs7025486	0.530229333
rs12078697	0.526071	rs11621859	0.535581667	rs7027205	0.544463667
rs1211375	0.529817333	rs11626974	0.525126	rs7038554	0.526510333

Appendix C

rs12129789	0.537832333	rs11664663	0.525008333	rs7045409	0.51757
rs12136856	0.515020333	rs11675880	0.532667333	rs7087685	0.534734667
rs12137505	0.528212333	rs1168330	0.526312667	rs7089306	0.513391
rs12138917	0.545657667	rs11687220	0.531572	rs7094463	0.513366333
rs12148239	0.523720667	rs11690007	0.525237333	rs7096909	0.528211667
rs12150079	0.522882333	rs11700764	0.536947667	rs7102454	0.525599667
rs12150672	0.520693	rs11713534	0.530038	rs7117121	0.520057333
rs12153375	0.542422333	rs11714281	0.537106	rs7124676	0.531706667
rs1216743	0.516253333	rs11734556	0.533553333	rs7125196	0.510949667
rs12204273	0.510205	rs11737660	0.523613	rs8045821	0.533661333
rs12210292	0.510636667	rs11752362	0.527762333	rs8057236	0.532261333
rs12246717	0.520381333	rs11757427	0.535773333	rs8060113	0.522809667
rs12258052	0.526877333	rs11763341	0.547166667	rs8075920	0.531752
rs12286037	0.513882667	rs11770655	0.518066333	rs8083385	0.528479333
rs12286929	0.514642667	rs1177290	0.526839667	rs8083614	0.532365
rs12288512	0.523815333	rs1177298	0.532283	rs8191425	0.524077
rs12290350	0.524310667	rs11852815	0.536124667	rs843205	0.527632667
rs12316047	0.54078	rs11876744	0.522794	rs848210	0.524460667
rs1233367	0.513619	rs11877115	0.523964	rs850763	0.524007333
rs12338788	0.533172	rs1187873	0.524157	rs852620	0.528696333
rs12402406	0.510853333	rs11915399	0.53447	rs865094	0.523955333
rs12406019	0.514191667	rs11917356	0.537501667	rs868641	0.487355667
rs12413409	0.511655333	rs11938521	0.533547333	rs872625	0.531448333
rs12420422	0.512014333	rs11966463	0.516717	rs875377	0.529317
rs1242229	0.519871333	rs11988085	0.528881667	rs882172	0.522542667
rs12430890	0.515235667	rs11997295	0.530440333	rs885561	0.524211333
rs12464308	0.525549667	rs1201559	0.513794	rs886527	0.532627333
rs12474201	0.527218	rs12026222	0.544112	rs887622	0.534867333
rs1249150	0.515194667	rs12029988	0.527476667	rs890448	0.538383667
rs12494849	0.537228	rs12085600	0.528445667	rs900791	0.53713
rs1250247	0.51887	rs12091336	0.536670333	rs907066	0.522839667
rs12517906	0.536845667	rs12131620	0.537677	rs912759	0.536271667
rs12518748	0.532512333	rs12147450	0.525437	rs9200	0.530824667
rs12537	0.539839	rs1227051	0.527470667	rs920559	0.531000333
rs1254260	0.519236667	rs1227634	0.527838333	rs921268	0.528331
rs12564600	0.514412667	rs1227635	0.541905667	rs921855	0.531415
rs12567136	0.517522333	rs12285610	0.532495667	rs922878	0.529388333
rs12571964	0.513293	rs12303068	0.521769	rs925111	0.532932667
rs12572775	0.528918	rs1233391	0.521608	rs9263715	0.531237
rs12575252	0.511840333	rs1233397	0.520873333	rs9268541	0.523156
rs12591128	0.519769667	rs12381471	0.523763667	rs9275765	0.543386
rs12597579	0.517705667	rs1239771	0.533915	rs6607337	0.51906
rs12604935	0.546653667	rs12414475	0.537117667	rs6656137	0.51199
rs12623218	0.518360667	rs1241482	0.524329	rs6657944	0.516904333
rs1263173	0.513474	rs12435594	0.541917333	rs6676084	0.545593667
rs12644	0.512336	rs12445870	0.539847333	rs6684770	0.522902
rs12654264	0.517053	rs12461127	0.533365	rs6688233	0.531029667

Appendix C

rs12654393	0.53362	rs12497339	0.528585667	rs6691053	0.524126
rs1265566	0.521983667	rs12509826	0.522045333	rs6694158	0.528374667
rs12677146	0.515038	rs12533343	0.528079667	rs6696923	0.527505
rs12696304	0.534258	rs12534073	0.537914	rs6704991	0.525715
rs12708980	0.527889	rs12542237	0.523805667	rs6708331	0.518511333
rs12712870	0.512528667	rs12542851	0.535071667	rs6711012	0.516283
rs12714414	0.518357667	rs12570292	0.518205	rs6714929	0.522203
rs12828016	0.51964	rs12588764	0.536984667	rs6723509	0.516459
rs1283706	0.525294667	rs12593611	0.520461333	rs6728726	0.513455333
rs12879626	0.515023333	rs12613663	0.53242	rs6730148	0.518743333
rs12885458	0.518352	rs12613732	0.520309	rs6736093	0.511808
rs12903590	0.516021667	rs12615223	0.522118667	rs6740249	0.518097
rs12909648	0.510547	rs12622458	0.531534333	rs6741676	0.534895667
rs12910825	0.514702333	rs12637547	0.506537333	rs6758859	0.513557667
rs12917449	0.519123	rs12661742	0.551694	rs6759676	0.513168333
rs12922830	0.51823	rs12682020	0.530515333	rs6777317	0.510388
rs12931627	0.52155	rs12692973	0.549161	rs6781254	0.516862667
rs12939549	0.528359	rs12694847	0.525227333	rs6793667	0.514057
rs1294421	0.512689333	rs12700545	0.526614	rs679992	0.521164667
rs12972158	0.518231667	rs12708994	0.534894667	rs6803322	0.520323
rs12979	0.524565	rs12727814	0.555916667	rs6813195	0.517366667
rs12983082	0.524682333	rs12735613	0.542703	rs681761	0.518908667
rs12986742	0.519580667	rs12822139	0.530234333	rs6830062	0.510719
rs13071657	0.518178667	rs12873360	0.531719333	rs6845078	0.511098333
rs13076366	0.513770333	rs12894688	0.535119	rs684573	0.560050667
rs13094430	0.521964667	rs1291114	0.519473	rs6857769	0.515375333
rs13134800	0.516748	rs1291401	0.532701333	rs6875372	0.511965
rs13160161	0.514110333	rs12919270	0.521560667	rs6876198	0.531886667
rs13196792	0.523231	rs1295348	0.519642667	rs688	0.525228
rs13210323	0.536745667	rs12984247	0.533440333	rs7653460	0.524814
rs13230111	0.515017333	rs12996329	0.530603	rs7660368	0.530088
rs13251900	0.522110333	rs12999804	0.516779333	rs769149	0.526609333
rs132644	0.518424667	rs13028985	0.457949667	rs7694725	0.530931
rs13273088	0.521635	rs13061191	0.530803	rs7702763	0.538908667
rs13277801	0.519107	rs13072883	0.531751333	rs7711765	0.531195333
rs13299788	0.519774667	rs13087528	0.520085333	rs7742617	0.539962667
rs13320646	0.514628667	rs1309615	0.532363667	rs7745661	0.549154667
rs1333042	0.532445667	rs13119988	0.539969667	rs7756056	0.534522333
rs1333048	0.532803	rs13134467	0.532560333	rs7756362	0.536276667
rs1333049	0.533468667	rs13141994	0.538176667	rs7759860	0.522073667
rs13333226	0.522247	rs13153782	0.535305333	rs7784067	0.526970333
rs1335587	0.516104333	rs13188259	0.547565	rs7784712	0.522574
rs13429325	0.514453	rs13200022	0.534258	rs7785364	0.531890333
rs1344672	0.514187333	rs1323423	0.529654	rs7790654	0.535076333
rs1346841	0.520084333	rs1323530	0.533624	rs7796377	0.523376667
rs134871	0.518859667	rs13242357	0.529717333	rs7800092	0.524061333
rs1361831	0.528533333	rs1324886	0.530023	rs7864580	0.536001667

Appendix C

rs1365466	0.519574	rs13272100	0.542809333	rs789559	0.530726333
rs1368298	0.525327667	rs13273954	0.533297333	rs7903665	0.522823
rs1372878	0.519961	rs13281575	0.539238333	rs7922034	0.530149333
rs1374313	0.527169333	rs13388646	0.517773	rs7924473	0.528212667
rs1374326	0.520898	rs1345300	0.531587	rs7946375	0.535422
rs1383592	0.512515	rs1357818	0.541386667	rs7947780	0.526341333
rs1394879	0.513793333	rs1362811	0.542706333	rs796126	0.527246333
rs1395594	0.519405667	rs1364114	0.523384667	rs7963300	0.528564333
rs1409158	0.525125	rs1372876	0.515159	rs7975207	0.534705667
rs1415701	0.524814333	rs138780	0.542607	rs7987890	0.542032
rs1420341	0.529964667	rs1389661	0.528461333	rs7988462	0.560499333
rs1421203	0.511299333	rs1393492	0.530249	rs7994890	0.544614333
rs14234	0.544523	rs1395524	0.532542	rs7998181	0.533827333
rs1425482	0.516748	rs1398024	0.533496667	rs8004086	0.516152667
rs1437134	0.525554333	rs1400625	0.547448333	rs8014621	0.540200667
rs1441433	0.537750333	rs1404608	0.511693333	rs8023044	0.517221333
rs1442493	0.526962333	rs1407793	0.518144667	rs8029216	0.530821
rs1443926	0.526388333	rs1408986	0.524274	rs6432053	0.511970333
rs1451506	0.51309	rs1409254	0.530913	rs6433243	0.521038
rs1451538	0.517617	rs1409838	0.536512333	rs643531	0.515714667
rs1472721	0.543482	rs1412259	0.539090667	rs6436755	0.513291667
rs1473482	0.516813667	rs1416390	0.524078333	rs6438234	0.526530333
rs1474476	0.522657333	rs1421898	0.522283	rs6440003	0.51434
rs1484375	0.517609	rs1422257	0.519186333	rs644045	0.514539333
rs1484475	0.519413	rs1427087	0.560317667	rs6451675	0.527687333
rs1484803	0.513515667	rs1430155	0.531784333	rs6453139	0.512726
rs1489502	0.518129	rs1441747	0.518244333	rs6456498	0.513106333
rs1495377	0.521687	rs1442407	0.528860333	rs6457796	0.525225
rs1496479	0.516764333	rs1445059	0.540813333	rs6458501	0.517393333
rs1501672	0.515327333	rs1445846	0.538540333	rs6462079	0.520573
rs1501720	0.521048	rs1450913	0.554417667	rs6463489	0.517225667
rs1502337	0.519088667	rs1456631	0.523166333	rs6476030	0.526851333
rs1505307	0.537091667	rs1456633	0.522656333	rs6477694	0.519016
rs1528450	0.538651333	rs1458198	0.531440333	rs6478453	0.525366667
rs1530530	0.512205333	rs1463388	0.546046	rs6482729	0.529803667
rs1534420	0.510773	rs1468177	0.539493	rs6489721	0.534747
rs1534696	0.515543	rs1470209	0.538448667	rs6494005	0.514239
rs1539172	0.517053	rs1471759	0.541599667	rs6497373	0.514022333
rs1547247	0.513449667	rs1479273	0.533073667	rs6506429	0.523170667
rs1554790	0.513785333	rs1480617	0.538817	rs6506897	0.542870667
rs1556562	0.517677667	rs1482047	0.520868333	rs651862	0.510640667
rs1562445	0.526417	rs1482577	0.530579667	rs652722	0.519163333
rs1571217	0.525453667	rs1483067	0.526102333	rs6540549	0.518943
rs1582114	0.515324667	rs14976	0.523228	rs6541229	0.522022333
rs161792	0.523138667	rs1500463	0.526208333	rs6548238	0.511106333
rs1620668	0.511637	rs1516807	0.548994	rs6551716	0.515530667
rs162366	0.514476	rs1518681	0.526794667	rs6558173	0.528848

Appendix C

rs1624134	0.511415	rs1521879	0.536303	rs6569648	0.511523
rs1624802	0.525204333	rs1523320	0.534431	rs6587552	0.519866333
rs1631619	0.514396667	rs1528472	0.534239	rs6587553	0.529210667
rs163818	0.511071667	rs1541855	0.52952	rs6589567	0.514618
rs16827981	0.517469333	rs1549759	0.527146	rs6601465	0.526237667
rs16832246	0.524126667	rs1551872	0.533815	rs7319335	0.543955333
rs16843947	0.514558	rs1566889	0.525290333	rs732109	0.536598667
rs16845301	0.513628667	rs1567986	0.529650333	rs7326447	0.525646667
rs16846100	0.520101	rs1572341	0.532905	rs7328706	0.526967333
rs16847402	0.515641667	rs1580239	0.520268333	rs733099	0.53676
rs16858462	0.511670667	rs1596996	0.525251667	rs7335152	0.542497
rs1688040	0.510665	rs1609459	0.525317	rs7394810	0.534336
rs16892892	0.514527333	rs1610247	0.523784667	rs739710	0.519295
rs16893526	0.520213667	rs1643889	0.516690333	rs741195	0.522842
rs1689437	0.534695333	rs164989	0.537172	rs743446	0.524671333
rs16903232	0.511378667	rs1662172	0.529236333	rs7462138	0.544466667
rs16917117	0.514276667	rs166296	0.524893	rs7486001	0.517382
rs16917237	0.523009	rs16822477	0.523843	rs7500315	0.544841667
rs16931177	0.527463	rs16825658	0.516722333	rs7516534	0.523386
rs1693551	0.520724	rs16827796	0.527872333	rs7553222	0.551351333
rs16940170	0.519977333	rs16834616	0.534984667	rs755495	0.535729667
rs16945088	0.516816	rs16838664	0.536181667	rs7555470	0.535021667
rs16952362	0.543600333	rs16852987	0.524621667	rs756550	0.524286667
rs16953467	0.515787667	rs16855560	0.518765667	rs7566764	0.530497667
rs16961543	0.532174333	rs16859886	0.528276667	rs75795	0.526636667
rs16962399	0.511426667	rs16877062	0.519008667	rs7580494	0.533267667
rs1696757	0.512191667	rs16888468	0.533086333	rs7584187	0.547416
rs16973236	0.510809667	rs16889280	0.535103	rs7585061	0.537175667
rs16988333	0.51849	rs16897054	0.53708	rs758658	0.550238
rs16996148	0.512901667	rs16899524	0.528461667	rs7587816	0.520227333
rs16996700	0.519806667	rs16927458	0.496701333	rs7602715	0.531315667
rs1701829	0.5142	rs16928081	0.522277667	rs7608731	0.517072667
rs17025214	0.510249333	rs16938799	0.520861667	rs7618619	0.526816667
rs17035270	0.518542667	rs16950362	0.532819333	rs7620870	0.522926333
rs17036328	0.516251	rs16958564	0.530157667	rs7623233	0.52453
rs17049741	0.513896333	rs1695960	0.531071333	rs7624419	0.538635
rs17053082	0.523276	rs1695969	0.535799667	rs7624766	0.528994667
rs17094182	0.510923667	rs16963190	0.522268333	rs763317	0.511922333
rs17101521	0.515836	rs16971976	0.529229333	rs7636752	0.520046
rs17119370	0.526775667	rs16980096	0.520695333	rs763780	0.518964333
rs17122278	0.525874667	rs17000816	0.522513	rs571411	0.515086
rs17178527	0.532083667	rs17007920	0.539538333	rs5749446	0.511728667
rs17242380	0.518795333	rs17007931	0.539524667	rs575030	0.517523667
rs17249754	0.525137	rs17008119	0.521528667	rs5751239	0.529200333
rs172629	0.522028	rs17009220	0.519281667	rs5754701	0.518101
rs17264866	0.521679	rs17010833	0.520974	rs5762311	0.513329
rs17275724	0.530724667	rs17011081	0.527507667	rs5762839	0.518987

Appendix C

rs17276464	0.511091	rs17027257	0.527720333	rs5766691	0.530569
rs17278665	0.517514667	rs17038182	0.547119667	rs580809	0.52505
rs17291650	0.510751333	rs17038218	0.534860667	rs588641	0.514141333
rs17313232	0.519228	rs17039902	0.526875	rs588837	0.528168667
rs17326656	0.511567	rs17053102	0.532533333	rs6001834	0.531509333
rs17353624	0.51674	rs1706631	0.522796333	rs6006426	0.532603667
rs17367504	0.515785667	rs17067528	0.523603667	rs6019483	0.526165667
rs17369163	0.511708	rs17071007	0.530367333	rs6021247	0.524314
rs17370725	0.516951667	rs17075173	0.531615	rs6021276	0.514924
rs17394429	0.522968	rs17076328	0.538765667	rs6023649	0.51929
rs17396055	0.524846333	rs17077300	0.5246	rs6025590	0.542453333
rs17398575	0.519998	rs17080149	0.529777333	rs6032919	0.529277333
rs17405722	0.513193667	rs17089264	0.520991667	rs6036804	0.522629667
rs17410962	0.513336	rs17133183	0.540032333	rs6054471	0.522800667
rs17411024	0.525508	rs17150118	0.527112333	rs6056891	0.522494667
rs17411031	0.514074667	rs17154350	0.540057	rs6064117	0.513026
rs17411126	0.512286667	rs1715789	0.530368667	rs6073143	0.515315
rs17422122	0.522015667	rs17164532	0.521335	rs6078093	0.522968667
rs17440913	0.526958333	rs17166876	0.52443	rs609018	0.5185
rs17446091	0.515422	rs17176528	0.530341667	rs6092477	0.526276
rs174534	0.533238	rs17177133	0.526774667	rs615031	0.539474333
rs17454517	0.536074	rs17177912	0.513536667	rs615672	0.517244333
rs174547	0.536710667	rs17188807	0.538439	rs622513	0.515784333
rs174548	0.527272	rs17197320	0.51864	rs623479	0.538464667
rs174549	0.532118	rs17204569	0.525456	rs6265	0.526114333
rs17482753	0.516946	rs17236525	0.532547	rs6271	0.512631667
rs17489268	0.511968667	rs17236749	0.532732	rs6428739	0.512032
rs17506606	0.528682	rs17236770	0.525814333	rs642961	0.531045
rs17561351	0.511422667	rs17241889	0.52413	rs7014346	0.527693333
rs17568628	0.514586667	rs17243903	0.538847667	rs7020283	0.529151
rs175804	0.510459	rs17263985	0.520089667	rs702639	0.511500333
rs1761450	0.518173667	rs17305779	0.532446333	rs7031458	0.528505667
rs17616661	0.514801667	rs17340893	0.519721	rs706439	0.516624333
rs17624303	0.513677667	rs1734343	0.523956333	rs7069053	0.517972333
rs17638167	0.516535	rs1734422	0.530987667	rs7071206	0.549476667
rs17640714	0.514145	rs17368670	0.526405	rs7071884	0.539345333
rs17647863	0.51947	rs17441830	0.521958667	rs7072877	0.538962333
rs17651892	0.5268	rs17553775	0.528285333	rs7080350	0.529369667
rs17664900	0.519275667	rs17561583	0.528429333	rs7081521	0.535267333
rs1767141	0.512137	rs17567105	0.524078333	rs7081687	0.536285
rs17672135	0.512933	rs17570453	0.531479667	rs7088720	0.531130667
rs17688434	0.511334333	rs17578280	0.53304	rs712779	0.523377333
rs17691584	0.511902333	rs17583120	0.501394	rs7128566	0.526849333
rs17695092	0.512619333	rs17629104	0.526316667	rs7144401	0.524971333
rs17700028	0.511553	rs17645681	0.524879667	rs7155872	0.521450667
rs17703883	0.536382	rs17654475	0.523445333	rs7166692	0.524557667
rs17712208	0.520302333	rs17664267	0.529988	rs717344	0.530700667

Appendix C

rs17735265	0.515928	rs17667613	0.527313333	rs7178742	0.526759
rs17759796	0.519934333	rs176713	0.527809333	rs7198295	0.523086
rs17766335	0.526989333	rs17672688	0.524055333	rs7204945	0.532902
rs17776482	0.523271667	rs17685928	0.525417	rs7209945	0.538334333
rs17789218	0.518909	rs17749450	0.519688	rs7213829	0.525865667
rs17799476	0.537854	rs17758247	0.525278667	rs7236339	0.519754
rs17815608	0.515815667	rs17786532	0.522701333	rs7239176	0.522316333
rs17828399	0.510894	rs17792215	0.52838	rs7239883	0.506054667
rs179131	0.523452667	rs17797652	0.527418667	rs7245949	0.526089
rs1799898	0.534093	rs17813827	0.523524333	rs7253270	0.525168
rs1800775	0.525770667	rs1781931	0.536835667	rs7255412	0.528884333
rs1801282	0.516265333	rs17829676	0.533045333	rs728171	0.529863333
rs1809498	0.519010333	rs1809921	0.547568667	rs7296677	0.564564333
rs1822438	0.520647	rs1828819	0.534350333	rs7298857	0.533966
rs1840969	0.528539667	rs184159	0.528699	rs7307677	0.528414333
rs1847018	0.512795667	rs1859308	0.528144333	rs7311068	0.538707
rs1848797	0.513847333	rs1863037	0.536769	rs4899280	0.545679
rs1859168	0.548322667	rs1865826	0.531045333	rs4909609	0.514847
rs1866956	0.515052	rs1868538	0.528662333	rs4916749	0.511601
rs1867073	0.531680333	rs1873026	0.531486333	rs4918796	0.529745667
rs1871045	0.514750333	rs1874898	0.523528333	rs4923460	0.521120667
rs1891110	0.527098333	rs1875675	0.543742333	rs4926853	0.514975
rs1895941	0.544224333	rs1887054	0.526573667	rs4930352	0.524496667
rs1899689	0.517158667	rs1890843	0.527029	rs4949465	0.521465667
rs190759	0.511285	rs1891067	0.522707333	rs495348	0.519899667
rs1912631	0.537285333	rs1907971	0.533295333	rs4953618	0.514653667
rs1919484	0.511015667	rs1909475	0.538577333	rs4959431	0.518163333
rs1931990	0.513498333	rs1909917	0.537546	rs496321	0.518568
rs1936797	0.526701667	rs1913477	0.523697667	rs4969183	0.519885667
rs193968	0.532132333	rs1916309	0.515180333	rs4970705	0.520469
rs1941706	0.528028	rs1927464	0.536180667	rs4970712	0.518151667
rs1943226	0.513195667	rs1929734	0.530729333	rs4977574	0.535447667
rs1967115	0.537094	rs1931565	0.521572333	rs4987358	0.5367
rs197374	0.530499667	rs1947551	0.523974333	rs515071	0.531844667
rs1976053	0.528812	rs1958375	0.528952	rs518295	0.532421
rs1991705	0.517834667	rs1971518	0.529651	rs518785	0.518554333
rs1991866	0.517137667	rs1981958	0.527987667	rs5215	0.525787
rs2001945	0.510687333	rs1994104	0.519453667	rs524492	0.541592
rs2007471	0.519837667	rs1995381	0.538352667	rs527248	0.529287333
rs2011442	0.539077333	rs1999078	0.534034333	rs527402	0.510697
rs2011603	0.52174	rs2007972	0.528924333	rs534523	0.514766667
rs201198	0.527107333	rs2030183	0.532864667	rs536338	0.53272
rs2029085	0.519930667	rs203202	0.541244667	rs541862	0.514450333
rs2048327	0.530743333	rs2035987	0.519048667	rs54211	0.526114333
rs2048508	0.513684667	rs2036098	0.53584	rs550677	0.513099
rs2049477	0.519032	rs2048428	0.530439	rs552525	0.520847
rs2051549	0.523462	rs2049661	0.533379667	rs558003	0.516661

Appendix C

rs2054903	0.523383	rs2053816	0.519577667	rs559390	0.526939667
rs2058619	0.514851667	rs2056962	0.534723	rs564398	0.519791333
rs2058919	0.527294333	rs2059776	0.539723667	rs564449	0.519508
rs2068063	0.517531333	rs2068114	0.532284	rs6764121	0.549639
rs2068834	0.513584667	rs2071918	0.525044667	rs6772329	0.520082
rs2070527	0.520573667	rs2076533	0.530789667	rs6776877	0.540210667
rs2070929	0.521868667	rs2077181	0.525849	rs6781428	0.511557
rs2072209	0.514255	rs2079753	0.553165333	rs6782002	0.525228667
rs2075803	0.529613667	rs2090587	0.530520333	rs6787443	0.527305667
rs2107713	0.525825667	rs2092741	0.524610333	rs6801593	0.533852667
rs2108978	0.512384667	rs211221	0.531024667	rs6809018	0.543619667
rs2112783	0.514337	rs2115435	0.531153	rs6816625	0.532496333
rs2131941	0.511102	rs2118322	0.531177	rs6821770	0.53547
rs2134964	0.520382	rs2119433	0.533039333	rs6824779	0.523321667
rs2146641	0.517888667	rs2139454	0.531936333	rs6841127	0.531202333
rs2156250	0.520960667	rs2149830	0.526543333	rs6846226	0.527515333
rs215634	0.514626	rs2152076	0.527835	rs6852190	0.526875333
rs2158177	0.522437333	rs2157465	0.530516333	rs6857595	0.525739667
rs2161738	0.516375333	rs2160052	0.518957667	rs6859219	0.532086333
rs2168608	0.510829333	rs2178528	0.529721333	rs6860569	0.528968
rs2171690	0.511208	rs2190921	0.524062	rs686137	0.523837667
rs217381	0.515280667	rs2191500	0.524565	rs6865875	0.529718667
rs2188151	0.517667	rs2203805	0.524700667	rs687174	0.533672333
rs2190788	0.533287333	rs2207729	0.522177333	rs6877743	0.529618667
rs2198776	0.512879667	rs2216739	0.536420333	rs6913822	0.52323
rs2208589	0.519863667	rs2237813	0.530042333	rs6914153	0.539285
rs2210612	0.545321	rs2238277	0.532980667	rs6915880	0.532858333
rs2212450	0.515599	rs2250861	0.521990667	rs6917809	0.532220333
rs2216707	0.525247	rs2254638	0.523035	rs6927500	0.525128333
rs2219939	0.533992	rs2255071	0.529604333	rs6932387	0.536247
rs223413	0.516878667	rs2272810	0.525616	rs6942503	0.521841667
rs2235797	0.544286333	rs2277725	0.541506	rs6943538	0.530401
rs2236033	0.512133333	rs2289439	0.528184667	rs694444	0.532571
rs2251393	0.516391	rs2293370	0.547155333	rs6947660	0.527859667
rs2252641	0.522804333	rs2298613	0.534397333	rs698295	0.530271
rs2253310	0.529837667	rs2299269	0.529906667	rs701158	0.536748
rs2256332	0.523525333	rs2299776	0.528108667	rs7013304	0.530927
rs2269997	0.539022333	rs2302972	0.515823667	rs4682718	0.520967667
rs2271904	0.511802	rs2303879	0.502802	rs4691380	0.512079
rs2272007	0.517265	rs2304787	0.522298333	rs4701031	0.539072667
rs2273	0.512956	rs2305602	0.524644333	rs4705938	0.512999
rs2273368	0.520269333	rs2306213	0.524676333	rs4708899	0.522427
rs2274089	0.510507667	rs2306707	0.531460333	rs471364	0.532050333
rs2274465	0.530893667	rs2311984	0.528731333	rs4737009	0.518289
rs2282032	0.516791333	rs2312949	0.477488333	rs4743034	0.535800333
rs2286276	0.529802667	rs2326366	0.536417	rs4743038	0.540707667
rs2292893	0.556451667	rs2326972	0.528716333	rs4745062	0.516782667

Appendix C

rs2297987	0.531764667	rs2332457	0.532876	rs4749369	0.52495
rs2298624	0.526096667	rs2339024	0.523914	rs4765127	0.532752
rs2301573	0.511523333	rs2346709	0.523129	rs4765219	0.528806333
rs2303014	0.513783667	rs2354124	0.517314	rs4771651	0.526417333
rs2321882	0.514287333	rs2360982	0.524100667	rs4774498	0.520777333
rs2335418	0.510590667	rs23840	0.515257667	rs4775031	0.516636333
rs2336230	0.511705333	rs2395656	0.528064	rs4775041	0.522564333
rs2338104	0.525483333	rs2397146	0.526060667	rs4775609	0.516635333
rs2345485	0.518002667	rs2399109	0.526386333	rs4783689	0.511631
rs2347252	0.532319333	rs2400963	0.532377333	rs4784227	0.513013667
rs2365716	0.513036667	rs2402168	0.529438333	rs4794047	0.515305333
rs2372888	0.531249	rs2411405	0.540172667	rs4795400	0.511422333
rs2373011	0.516299	rs2414817	0.537305	rs4800401	0.529849667
rs2373115	0.517809	rs2417862	0.532490333	rs4814512	0.516097667
rs2391518	0.525743667	rs2419879	0.526751667	rs4815191	0.515916333
rs2395607	0.511926667	rs2421005	0.534581333	rs4820079	0.520253667
rs241033	0.5211	rs2422008	0.532742667	rs4834232	0.517740333
rs241472	0.517635	rs2423646	0.515156667	rs4836339	0.515351
rs2415209	0.519225333	rs2431617	0.522903667	rs4841436	0.527198667
rs2425024	0.525215667	rs243849	0.522205	rs4846905	0.517520667
rs2434612	0.516741667	rs2440807	0.526771333	rs4858697	0.518196333
rs2436728	0.523422333	rs2451233	0.536638667	rs4860782	0.521296667
rs2439823	0.511916	rs2453605	0.539034667	rs487152	0.528616333
rs244415	0.522727667	rs2489394	0.531433	rs4881702	0.533287
rs2444820	0.516572333	rs2522422	0.525258667	rs4886655	0.519826
rs2469997	0.524870333	rs2523485	0.521429	rs4890633	0.525516667
rs2479409	0.52498	rs2523691	0.525256333	rs648210	0.530951
rs2482506	0.527842	rs2527894	0.524600333	rs648460	0.535006333
rs2486942	0.513080667	rs2564118	0.530250333	rs6493489	0.557942
rs2487928	0.511368333	rs2570501	0.529962333	rs6495768	0.523864333
rs2504236	0.510529667	rs2609839	0.528058667	rs6496669	0.521652
rs2516739	0.532616	rs2642780	0.537471333	rs6497035	0.530083
rs2543029	0.512382333	rs2651927	0.529145	rs6499766	0.534584
rs2549794	0.511069667	rs26680	0.534704	rs6501798	0.534501667
rs255900	0.512633333	rs26737	0.515843	rs6536287	0.521833
rs2562784	0.562953	rs2683120	0.526985667	rs654128	0.544072333
rs2573566	0.510856	rs2687661	0.532862	rs6547082	0.528242333
rs2575570	0.516151333	rs2716081	0.521480667	rs6549009	0.538140667
rs2596517	0.518552667	rs2758912	0.534537667	rs6560608	0.526461333
rs2605082	0.511197	rs2769896	0.536458	rs6564635	0.552326333
rs2607426	0.522225333	rs2792954	0.49323	rs6573038	0.525081667
rs261076	0.534145	rs2796050	0.519630333	rs6574512	0.526300667
rs2611742	0.524157667	rs2817308	0.526157333	rs6575434	0.518108667
rs2615453	0.51829	rs2829355	0.520303333	rs657972	0.541673667
rs262566	0.522357667	rs2829934	0.518854667	rs6579845	0.546179667
rs2627197	0.526937333	rs2830255	0.526456667	rs6585272	0.522500333
rs263169	0.536655333	rs283091	0.530253333	rs6589531	0.522252333

Appendix C

rs263173	0.535072333	rs2832075	0.530375	rs6594987	0.533579667
rs2633725	0.520177667	rs2832147	0.532132	rs6660584	0.530395333
rs2634047	0.529051667	rs2832191	0.537841	rs6682150	0.552737
rs2640017	0.518563	rs2836527	0.527472667	rs6685584	0.519141
rs2657879	0.515294667	rs283730	0.531672	rs6686886	0.526494667
rs266274	0.534269333	rs2837853	0.536047667	rs6690359	0.529846333
rs2665118	0.515787333	rs28394649	0.527967667	rs6695961	0.522629667
rs2689700	0.528359	rs2844043	0.545083	rs6708609	0.525916333
rs2692894	0.525348667	rs2846607	0.534529333	rs6714968	0.522809
rs2699808	0.519399667	rs2864755	0.523229	rs6724173	0.540106667
rs2708146	0.517592	rs2871987	0.536457333	rs6731889	0.521968333
rs2721954	0.524960333	rs2877260	0.544776333	rs6733471	0.526407333
rs2732159	0.522501667	rs288139	0.528838667	rs6736039	0.528204333
rs2737203	0.525324333	rs288193	0.528654	rs6737578	0.514624667
rs2742540	0.511924333	rs2883782	0.505456	rs6740906	0.536278667
rs2745349	0.513409667	rs2885	0.533585667	rs4379706	0.536617333
rs2745865	0.525365667	rs2898425	0.536314667	rs4385527	0.527736333
rs27687	0.515104	rs2907358	0.523656	rs4389656	0.512351
rs2783130	0.516291333	rs2912524	0.536146	rs4409766	0.518746667
rs278640	0.529574667	rs291640	0.534919333	rs4420537	0.515245
rs2807834	0.523295	rs2918194	0.523315333	rs4458918	0.546907
rs2812	0.520404333	rs2940305	0.519915333	rs4468878	0.516926
rs2814944	0.511436	rs2964798	0.524279333	rs448385	0.515479667
rs2814993	0.512968333	rs298528	0.522124333	rs4493170	0.513872667
rs2816177	0.510521667	rs3015729	0.531574667	rs4495442	0.513064667
rs281868	0.516127	rs303993	0.535888333	rs4506565	0.511003333
rs2820426	0.519766	rs3099849	0.516064667	rs4556142	0.513613
rs2823253	0.519152	rs3110606	0.515914	rs4568876	0.523836
rs2825652	0.514446667	rs3124994	0.531971667	rs4570167	0.523969
rs2832296	0.513384333	rs316762	0.542669	rs4582848	0.513902
rs2836633	0.516491	rs318572	0.542471333	rs4586544	0.513964
rs2838344	0.512177667	rs329655	0.520294333	rs4599004	0.510831
rs2838774	0.517243333	rs335419	0.533145	rs460976	0.513020667
rs28497362	0.526505667	rs343989	0.541588	rs4620037	0.515703333
rs285685	0.532556	rs34424	0.525102	rs4633	0.516046667
rs2861089	0.515401333	rs34602	0.526636667	rs4639796	0.529752333
rs2861699	0.528680667	rs352610	0.545642667	rs4646283	0.526362
rs2891168	0.535509333	rs352980	0.530346667	rs4646949	0.549633667
rs28927680	0.515124333	rs357233	0.554892333	rs4648633	0.524392
rs2899663	0.533198333	rs357833	0.544156667	rs4650994	0.518165333
rs2903492	0.516271	rs371925	0.528328667	rs4655772	0.518108667
rs2916577	0.527721667	rs3745469	0.528066667	rs4660293	0.521235333
rs2923437	0.519698333	rs3751527	0.539007	rs4672317	0.515121667
rs2941504	0.533942	rs3757674	0.520092667	rs4676084	0.526117
rs2942194	0.519316	rs3765339	0.531844	rs4679848	0.513979333
rs2959592	0.525321	rs3769825	0.527987	rs539853	0.530308667
rs2969	0.514447	rs3769827	0.533967	rs553998	0.517463333

Appendix C

rs2980862	0.512027333	rs3774108	0.528311	rs556739	0.530049667
rs3006917	0.515870667	rs3792267	0.539960667	rs560859	0.528543667
rs300934	0.521328333	rs3793177	0.532239667	rs571247	0.531930333
rs3010239	0.521267	rs3793408	0.516715333	rs5761499	0.535273
rs3010276	0.517792333	rs3796504	0.519428667	rs579005	0.549252
rs3087595	0.536548	rs3799731	0.531488667	rs5992985	0.532071667
rs3088050	0.516407	rs3803064	0.533427	rs6009503	0.521866667
rs3094061	0.510712	rs3816240	0.519500667	rs6027951	0.533504667
rs3096490	0.521772	rs3828783	0.536104333	rs603315	0.541473
rs3106598	0.521389667	rs3847888	0.542851333	rs6066361	0.534498
rs3116602	0.522158	rs3853299	0.550195333	rs6075709	0.541025667
rs3118914	0.522193667	rs386397	0.519809667	rs6076925	0.523763667
rs311904	0.517455333	rs3866555	0.54776	rs609048	0.52777
rs312750	0.517680667	rs3884325	0.539292333	rs609098	0.527025667
rs3129063	0.544304333	rs3886607	0.535068333	rs6101100	0.527765
rs312924	0.545824333	rs3886870	0.550998	rs6102825	0.528489333
rs3129768	0.514436	rs3897456	0.535453	rs611959	0.532168333
rs3130048	0.542446	rs391583	0.518762667	rs6129032	0.52671
rs3130501	0.518142	rs3918315	0.528342	rs6134506	0.519824667
rs3131294	0.513511333	rs392066	0.530307333	rs6140981	0.531274667
rs3132524	0.516924667	rs3924164	0.505794333	rs635578	0.520259333
rs3132718	0.520976	rs394307	0.530452333	rs639622	0.536740333
rs314253	0.521996333	rs3948464	0.539531333	rs6419833	0.531631
rs3176466	0.512174333	rs3957111	0.532799	rs6438213	0.537617
rs3198697	0.519209667	rs3997982	0.516593333	rs6444190	0.527993
rs3211931	0.520634333	rs404513	0.523959333	rs6463443	0.553658667
rs321358	0.524816667	rs404678	0.543299	rs6471656	0.531659333
rs3217805	0.528015	rs4073220	0.525945333	rs6477547	0.521574667
rs322699	0.524324333	rs41336646	0.54966	rs4803168	0.536175667
rs333947	0.519905	rs41342048	0.531227333	rs4805867	0.525553667
rs342989	0.523752333	rs413721	0.521611333	rs4809647	0.530245333
rs343000	0.522813667	rs41402445	0.544749333	rs4814980	0.521421667
rs346617	0.520723667	rs4143549	0.533782	rs4817258	0.524383667
rs348495	0.520875333	rs41462844	0.528917	rs4821901	0.539018667
rs35183060	0.512959333	rs41465748	0.535686	rs4843860	0.528381
rs355773	0.516944667	rs41483746	0.519804333	rs4844381	0.529738333
rs355909	0.510931667	rs41499949	0.522633333	rs4845898	0.529208333
rs360804	0.524439333	rs41522	0.535026	rs4851095	0.528303333
rs367943	0.516038333	rs422224	0.530841667	rs4851877	0.530892
rs368863	0.513569	rs4233678	0.521825	rs4857907	0.516785333
rs372558	0.536855667	rs4234222	0.520816333	rs4867597	0.531628333
rs3729639	0.510989	rs4235496	0.531196	rs4904019	0.524485667
rs3731544	0.510043333	rs4238250	0.529028333	rs4906742	0.538851
rs3732837	0.534099333	rs4270483	0.521442333	rs4908481	0.522648667
rs3734618	0.516659667	rs4301023	0.524874667	rs4908602	0.54139
rs3737136	0.519136	rs4301655	0.525009	rs4912036	0.525215333
rs3738815	0.524934333	rs4322586	0.537238	rs4917653	0.539145667

Appendix C

rs3739998	0.51831	rs4389718	0.534756333	rs4938445	0.546332
rs3740390	0.513485333	rs4439672	0.521230667	rs4940646	0.531832
rs3745651	0.511623333	rs4443226	0.524740667	rs4943543	0.529487
rs3746337	0.536017667	rs4445711	0.535533667	rs4944092	0.535546
rs3754507	0.516177	rs4473690	0.513832	rs4955239	0.525280667
rs3755652	0.522090333	rs4509055	0.527219	rs4968816	0.520300333
rs3759811	0.558916333	rs4521323	0.538342	rs4981148	0.536628667
rs3764835	0.520315333	rs4539953	0.536032667	rs5009910	0.516924
rs3766430	0.513948667	rs4549	0.549540667	rs535457	0.529033
rs3766612	0.536282333	rs4567005	0.526333	rs389546	0.527786333
rs3769873	0.512071333	rs4567983	0.535807	rs3905000	0.518871667
rs3773910	0.519011	rs4580841	0.532197667	rs3916765	0.522205
rs3774372	0.525030667	rs4587081	0.520952	rs39201	0.522513333
rs3775002	0.519405333	rs4597906	0.540277667	rs3934712	0.517936667
rs377763	0.527494	rs4627591	0.552674	rs3999089	0.513249333
rs3780181	0.510164	rs4646649	0.532631667	rs4003412	0.519165
rs3787186	0.514631333	rs4670704	0.524251333	rs40060	0.519482333
rs3790268	0.516288333	rs4684800	0.538844	rs4008004	0.517103
rs3808460	0.525947	rs4699215	0.539322	rs40136	0.51065
rs3808607	0.519259333	rs4720876	0.536104333	rs4044321	0.518181333
rs3809060	0.515084667	rs4722820	0.529955	rs405001	0.521118333
rs3809346	0.519030333	rs4727512	0.52781	rs4075958	0.518566
rs3814115	0.545956	rs4751101	0.524327	rs408302	0.528597
rs3818638	0.513705	rs4758510	0.527838333	rs41177	0.536239667
rs3825801	0.512496667	rs4763063	0.523198333	rs4128868	0.527803667
rs3843467	0.523319667	rs4765905	0.526115667	rs4131682	0.522644333
rs3846663	0.517900667	rs4780957	0.527291667	rs41360247	0.527069667
rs3847687	0.512301333	rs4783311	0.537553	rs41411047	0.511485667
rs3850422	0.524181667	rs4784708	0.51471	rs4149268	0.517880333
rs3853445	0.512737667	rs4788684	0.526308667	rs41524846	0.527822667
rs3865188	0.511942333	rs4791171	0.529994667	rs4242231	0.521766667
rs3890182	0.515289333	rs4801417	0.528887333	rs4269585	0.514262667
rs4343164	0.519355	rs4317857	0.511165333	rs4275659	0.516961
rs4345341	0.532529333	rs4327091	0.52104		

WHAT IS CLAIMED IS:

1. A kit for determining methylation status of at least one CpG dinucleotide and
5 a genotype of at least one single-nucleotide polymorphism (SNP), the kit comprising:
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 cg04988978, cg21161138,
cg12655112, cg03725309, cg12586707, and cg17901584 or at a second CpG dinucleotide in
10 linkage disequilibrium with the first CpG dinucleotide at a GC locus selected from the group
consisting of cg04988978, cg21161138, cg12655112, cg03725309, cg12586707, and
cg17901584, 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
15 complementary to a DNA sequence or a bisulfite-converted DNA sequence of a first SNP
selected from the group consisting of rs2869675, rs4376434, rs12129789, rs7585056,
rs710987, rs4639796, rs1333048, rs12714414, rs942317, and rs1441433 or a second SNP in
linkage disequilibrium with the first SNP selected from the group consisting of rs2869675,
rs4376434, rs12129789, rs7585056, rs710987, rs4639796, rs1333048, rs12714414,
20 rs942317, and rs1441433, wherein the linkage disequilibrium has a value of $R > 0.3$.
2. The kit of claim 1, wherein the at least one first nucleic acid primer detects the
unmethylated CpG dinucleotide.
- 25 3. The kit of claim 1, wherein the at least one first nucleic acid primer detects the
methylated CpG dinucleotide.
4. The kit of any of claims 1-3, wherein the at least one first nucleic acid primer
comprises one or more nucleotide analogs.

30

5. The kit of any of claims 1-4, wherein the at least one first nucleic acid primer comprises 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 the at least one first nucleic acid primer 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 in a biological sample from a subject, wherein the biomarkers are associated with detecting CVD, determining severity of CVD, estimating survival from CVD, identifying, customizing, and/or optimizing intervention(s) for CVD, managing CVD and/or monitoring CVD, the method 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 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 cg04988978, cg21161138, cg12655112, cg03725309, cg12586707, and cg17901584, or a second CpG dinucleotide in linkage disequilibrium with the first CpG dinucleotide at a GC locus selected from the group consisting of cg04988978, cg21161138, cg12655112, cg03725309, cg12586707, and cg17901584, 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 rs2869675, rs4376434, rs12129789, rs7585056, rs710987, rs4639796, rs1333048, rs12714414, rs942317, and rs1441433 or a second SNP in linkage disequilibrium with a first SNP selected from the group consisting of rs2869675, rs4376434, rs12129789, rs7585056, rs710987, rs4639796, rs1333048, rs12714414, rs942317, and rs1441433, wherein the linkage disequilibrium has a value of $R > 0.3$,

wherein the percentage of methylation of the CpG dinucleotide at the GC locus selected from the group consisting of cg04988978, cg21161138, cg12655112, cg03725309, cg12586707, and cg17901584, and the identity of the nucleotide at the first SNP selected from the group consisting of rs2869675, rs4376434, rs12129789, rs7585056, rs710987, rs4639796, rs1333048, rs12714414, rs942317, and rs1441433 or the second SNP in linkage disequilibrium with the first SNP are biomarkers associated with detecting CVD or estimating survival from 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 the at least one first nucleic acid primer detects the unmethylated CpG dinucleotide.

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

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

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

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

10 19. A method of determining the presence of biomarkers in a biological sample from a subject, wherein the biomarkers are associated with detecting CVD, determining severity of CVD, estimating survival from CVD, identifying, customizing, and/or optimizing intervention(s) for CVD, managing CVD and/or monitoring CVD, the method comprising:

(a) obtaining a nucleic acid sample from the subject sample;

15 (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 selected from rs2869675, rs4376434, rs12129789, rs7585056, rs710987, rs4639796, rs1333048, rs12714414, rs942317, and rs1441433 or from Appendix C and/or is a second SNP in linkage disequilibrium ($R > 0.3$) with a first SNP selected from rs2869675, rs4376434, rs12129789, rs7585056, rs710987, rs4639796, rs1333048, rs12714414, rs942317, and
20 rs1441433 or 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 a second portion of the nucleic acid sample to detect methylation status of at least one CpG site selected from cg04988978, cg21161138, cg12655112, cg03725309, cg12586707, and cg17901584 or from Appendix A and/or a CpG
25 site collinear ($R > 0.3$) with a CpG selected from cg04988978, cg21161138, cg12655112, cg03725309, cg12586707, and cg17901584 or from Appendix A to obtain methylation data; and

(d) entering the genotype data from step (b) and/or methylation data from step
30 (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.

20. The method of claim 19, wherein 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.

21. The method of claim 19, wherein the at least one interaction effect is a gene-environment interaction effect (SNPxCpG) between a CpG site selected from cg04988978, cg21161138, cg12655112, cg03725309, cg12586707, and cg17901584 or from Appendix A or a CpG site that is collinear ($R > 0.3$) with a CpG site selected from cg04988978, cg21161138, cg12655112, cg03725309, cg12586707, and cg17901584 or from Appendix A and a SNP selected from rs2869675, rs4376434, rs12129789, rs7585056, rs710987, rs4639796, rs1333048, rs12714414, rs942317, and rs1441433 or from Appendix C or a SNP within moderate linkage disequilibrium ($R > 0.3$) from a SNP selected from rs2869675, rs4376434, rs12129789, rs7585056, rs710987, rs4639796, rs1333048, rs12714414, rs942317, and rs1441433 or from Appendix C.

22. The method of claim 19, wherein the at least one interaction effect is an environment-environment interaction effect (CpGxCpG) between at least two CpG sites selected from cg04988978, cg21161138, cg12655112, cg03725309, cg12586707, and cg17901584 or from Appendix A.

23. The method of claim 22, wherein 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 selected from cg04988978, cg21161138, cg12655112, cg03725309, cg12586707, and cg17901584 or from Appendix A.

24. The method of claim 19, wherein the at least one interaction effect is a gene-gene interaction effect (SNPxSNP) between at least two SNPs selected from rs2869675,

rs4376434, rs12129789, rs7585056, rs710987, rs4639796, rs1333048, rs12714414, rs942317, and rs1441433 or from Appendix C.

25. The method of claim 24, wherein one or both of the at least two SNPs are
5 collinear ($R > 0.3$) with one or both of the at least two SNPs selected from rs2869675, rs4376434, rs12129789, rs7585056, rs710987, rs4639796, rs1333048, rs12714414, rs942317, and rs1441433 or from Appendix C.

26. The method of any of claims 19-25, wherein the biological sample is a saliva
10 sample.

27. A system for determining methylation status of at least one CpG dinucleotide and a genotype of at least one single-nucleotide polymorphism (SNP), the system comprising:

15 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 selected from rs2869675, rs4376434, rs12129789, rs7585056, rs710987, rs4639796, rs1333048, rs12714414, rs942317, and rs1441433 or from Appendix C
20 and/or is a second SNP in linkage disequilibrium ($R > 0.3$) with a first SNP selected from rs2869675, rs4376434, rs12129789, rs7585056, rs710987, rs4639796, rs1333048, rs12714414, rs942317, and rs1441433 or from Appendix C to obtain genotype data;

a methylation assay module configured to bisulfite convert the nucleic acid in
25 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 selected from cg04988978, cg21161138, cg12655112, cg03725309, cg12586707, and cg17901584 or from Appendix A and/or a CpG site collinear ($R > 0.3$) with a CpG site selected from cg04988978, cg21161138, cg12655112, cg03725309, cg12586707, and
30 cg17901584 or 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 and/or methylation data.

28. The system of claim 27, wherein the algorithm is a machine learning
5 algorithm capable of accounting for linear and non-linear effects.

29. The system of claim 27 or 28, further comprising 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
10 main effect and/or at least one interaction effect based on the genotype data from step (b)
and/or methylation data from step (c).

30. A non-transitory computer-readable medium storing instructions executable
by a processing device to perform operations comprising:
15 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:

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,
20 wherein the at least one SNP is a first SNP selected from rs2869675, rs4376434, rs12129789,
rs7585056, rs710987, rs4639796, rs1333048, rs12714414, rs942317, and rs1441433 or from
Appendix C and/or is a second SNP in linkage disequilibrium ($R > 0.3$) with a first SNP
selected from rs2869675, rs4376434, rs12129789, rs7585056, rs710987, rs4639796,
rs1333048, rs12714414, rs942317, and rs1441433 or from Appendix C to obtain the
25 genotype data; and

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 selected from cg04988978, cg21161138, cg12655112,
cg03725309, cg12586707, and cg17901584 or from Appendix A and/or a CpG site collinear
30 ($R > 0.3$) with a CpG site selected from cg04988978, cg21161138, cg12655112, cg03725309,
cg12586707, and cg17901584 or from Appendix A to obtain methylation data.

31. The non-transitory computer-readable medium of claim 30, wherein the operations further comprise providing an output based on the accounting.

5 32. The non-transitory computer-readable medium of claim 31, wherein the output comprises 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
10 triggering the delivery of a therapeutic based on the accounting.

33. A kit for determining methylation status of at least one CpG dinucleotide, the kit comprising:

 at least one first nucleic acid primer at least 8 nucleotides in length that is
15 complementary to a bisulfite-converted nucleic acid sequence comprising a first CpG dinucleotide at a GC locus selected from the group consisting of cg04988978, cg21161138, cg12655112, cg03725309, cg12586707, and cg17901584 or at a second CpG dinucleotide in linkage disequilibrium with the first CpG dinucleotide at a GC locus selected from the group consisting of cg04988978, cg21161138, cg12655112, cg03725309, cg12586707, and
20 cg17901584, 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.

34. A method of determining the presence of biomarkers in a biological sample from a subject, wherein the biomarkers are associated with detecting CVD, determining
25 severity of CVD, estimating survival from CVD, identifying, customizing, and/or optimizing intervention(s) for CVD, managing CVD and/or monitoring CVD, the method comprising:

 (a) providing a biological sample from the subject at risk for or having CVD, wherein nucleic acids from at least a portion of the biological sample are bisulfite converted;
and

30 (b) contacting the bisulfite converted nucleic acids 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 cg04988978, cg21161138, cg12655112, cg03725309, cg12586707, and cg17901584, or a second CpG dinucleotide in linkage disequilibrium with the first CpG dinucleotide at a GC locus selected from the group consisting of cg04988978, cg21161138, cg12655112, cg03725309, cg12586707, and cg17901584, wherein the linkage disequilibrium has a value of $R > 0.3$, wherein the first nucleic acid primer detects a methylated or unmethylated CpG dinucleotide, wherein the percentage of methylation of the CpG dinucleotide at the GC locus selected from the group consisting of cg04988978, cg21161138, cg12655112, cg03725309, cg12586707, and cg17901584 is associated with estimating survival of the subject.

35. A method of determining the presence of biomarkers in a biological sample from a subject, wherein the biomarkers are associated with detecting CVD, determining severity of CVD, estimating survival from CVD, identifying, customizing, and/or optimizing intervention(s) for CVD, managing CVD and/or monitoring CVD, the method comprising:

- (a) isolating nucleic acid sample from the subject sample;
- (b) bisulfite converting at least a portion of the nucleic acid and performing methylation assessment on the bisulfite converted nucleic acid to determine the methylation status of at least one CpG site selected from cg04988978, cg21161138, cg12655112, cg03725309, cg12586707, and cg17901584 or from Appendix A and/or a CpG site collinear ($R > 0.3$) with a CpG site selected from cg04988978, cg21161138, cg12655112, cg03725309, cg12586707, and cg17901584 or from Appendix A to obtain methylation data; and
- (c) entering the methylation data from step (b) into an algorithm that accounts for at least one CpG main effect, wherein the algorithm is a machine learning algorithm capable of accounting for linear and non-linear effects.

36. A system for determining methylation status of at least one CpG dinucleotide, the system comprising:

- a nucleic acid isolation module configured to isolate a nucleic acid sample from a subject sample;
- a methylation assay module configured to bisulfite convert the nucleic acid in at least a portion of the nucleic acid and perform a methylation assessment on the bisulfite

converted nucleic acid to determine methylation status of at least one CpG site selected from cg04988978, cg21161138, cg12655112, cg03725309, cg12586707, and cg17901584 or from Appendix A and/or a CpG site collinear ($R > 0.3$) with a CpG site selected from cg04988978, cg21161138, cg12655112, cg03725309, cg12586707, and cg17901584 or from Appendix A to obtain methylation data; and

an identification system configured to account for at least one CpG main effect based on the methylation data.

37. A non-transitory computer-readable medium storing instructions executable by a processing device to perform operations comprising:

accounting for at least one CpG main effect based on methylation data, wherein:

the methylation data is based on a methylation assay on a bisulfite converted nucleic acid in at least a portion of a nucleic acid sample to detect methylation status of at least one CpG site selected from cg04988978, cg21161138, cg12655112, cg03725309, cg12586707, and cg17901584 or from Appendix A and/or a CpG site collinear ($R > 0.3$) with a CpG site selected from cg04988978, cg21161138, cg12655112, cg03725309, cg12586707, and cg17901584 or from Appendix A to obtain methylation data.