

1 DNA Methylation Predicts Adverse Outcomes of Coronary

2 Artery Disease

3

4 Min Qin^{1,2,3,8}, Qili Wu^{1,2}, Xiaoxue Tian^{1,2}, Qian Zhu^{1,2}, Xianhong Fang⁴, Xiaoping

5 Chen⁵, Chen Liu⁶, Bin Zhang⁴, Hanping Li², Xipei Wang⁷, Cuiping Pan^{8,9#}, Shilong

6 Zhong^{1,2,3,7#}

7

8 Affiliations:

9 1. Department of Pharmacy, Guangdong Provincial People's Hospital (Guangdong
10 Academy of Medical Sciences), Southern Medical University, Guangzhou, China

11 2. Guangdong Provincial Key Laboratory of Coronary Heart Disease Prevention,
12 Guangdong Cardiovascular Institute, Guangdong Provincial People's Hospital
13 (Guangdong Academy of Medical Sciences), Southern Medical University, Guangzhou,
14 China

15 3. School of Medicine, South China University of Technology, Guangzhou, China

16 4. Department of Cardiology, Guangdong Provincial Key Laboratory of Coronary Heart
17 Disease Prevention, Guangdong Cardiovascular Institute, Guangdong Provincial
18 People's Hospital (Guangdong Academy of Medical Sciences), Southern Medical
19 University, Guangzhou, China

20 5. Department of Clinical Pharmacology, Xiangya Hospital, Central South University,
21 Changsha, 410008, Hunan, China.

22 6. Department of Cardiology, The First Affiliated Hospital of Sun Yat-Sen University,
23 Guangzhou, 510080, Guangdong, China.

24 7. Laboratory of Phase I Clinical Trials, Medical Research Center of Guangdong
25 Provincial People's Hospital (Guangdong Academy of Medical Sciences), Southern
26 Medical University, Guangzhou, China

- 1 8. Center for Intelligent Medicine Research, Greater Bay Area Institute of Precision
- 2 Medicine (Guangzhou), School of Life Sciences, Fudan University, Guangzhou, China
- 3 9. Center for Evolutionary Biology, Fudan University, Shanghai, China.

4 **# Correspondence**

5 Cuiping Pan: pancuiping@ipm-gba.org.cn

6 Shilong Zhong: gdph_zhongsl@gd.gov.cn and zhongsl@hotmail.com

7

8

9

1	Table of Contents	
2	Abstract	4
3	Introduction	5
4	Methods	6
5	Cohort assembly and baseline information collection.....	6
6	DNA extraction from blood leukocytes.....	7
7	Genome-wide DNA methylation profiling and data preprocessing	8
8	Epigenome-wide association analysis.....	9
9	Characterizing genomic locations of DMPs	10
10	Target gene predictions.....	10
11	Construction of prognostic models for death and MACE.....	11
12	Pleiotropic association analysis of DMPs and eQTL.....	12
13	Genome-wide association study of death and MACE.....	13
14	Differential gene expression analysis in MI and stroke.....	13
15	Statistical tests.....	14
16	Results	14
17	Baseline characteristics of the study population	14
18	EWAS identified differentially methylated CpGs for death and MACE	15
19	Pathways and mediating phenotypes of DMPs	17
20	Prognostic models for death and MACE in CAD.....	19
21	Contribution of genetic regulation to CAD prognosis.....	21
22	Prognostic genes in acute MI and stroke	23
23	Discussion	25
24	Acknowledgements	29
25	Funding	29
26	Conflict of interest	29
27	Data availability	30
28	Code availability	Error! Bookmark not defined.
29	Author Contributions	30
30	Ethical approval	31
31	References	31
32	Figure Legends	37
33		
34		

1 **Abstract**

2 **Background**

3 Adverse outcomes including myocardial infarction and stroke render coronary artery
4 disease (CAD) a leading cause of death worldwide. Biomarkers that predict such
5 adversity enable closer medical supervision and opportunities for improved outcomes.

6 **Methods and results**

7 We present a study of genome-wide DNA methylation profiling in 933 CAD patients
8 with up to 13 years of clinical follow-up. We discovered 115 methylation sites
9 associated with poor prognosis and inferred that cellular senescence, inflammation,
10 and high-density lipoprotein mediated the adversity. We built succinct prognostic
11 models combining a few methylation sites and clinical features, which could stratify
12 patients of different risks. Furthermore, we assessed genetic regulation of the
13 differential methylation by interrogating QTL effects. Prognostic genes such as
14 *FKBP5*, *UBE2E2* and *AUTS2* appeared recurrently in various analyses and were
15 validated in patients of myocardial infarction and stroke.

16 **Conclusions**

17 Our study provides prognostic models for clinical application and revealed
18 methylation biomarkers and mechanisms of CAD adverse outcomes.

19 **Key words:**

20 coronary artery disease; prognostic model; machine learning; DNA methylation;
21 inflammation; death;

22

1 **Introduction**

2 Coronary artery disease (CAD) is life-threatening and represents a universal leading
3 cause of death ¹⁻³. Studies of the last century suggested a 15-year survival rate of 48-
4 70% ^{4,5}. Despite the remarkable amelioration in the recent 30 years in managing its
5 clinical risk factors and the secondary and tertiary preventions, CAD is associated
6 with 17.8 million annual deaths worldwide ^{6,7}. Identifying patients with greater risk of
7 poor prognosis enables closer medical supervision and therefore opportunities for
8 better clinical outcomes. Numerous genetics-based research reported novel targets
9 and tools for predicting adverse outcomes in CAD patients. Indeed, CAD has an
10 estimated heritability of 0.38-0.66 for incidence ⁸ and 0.38-0.57 for mortality ⁹.
11 However, towards which direction it progresses is multifactorial determined by the
12 combined effects of genetic and environmental factors, therefore we reason that
13 considering multiple layers of information, such as genetics and epigenetics, will
14 better identify patients susceptible to poor prognostic outcomes.

15 DNA methylation on CpG (cytosine-phosphate-guanine) dinucleotides, a
16 stable yet dynamic regulation mechanism reflecting both genetics and environment,
17 enables exploring their integrated effects on diseases. Epigenome-wide association
18 studies (EWAS) suggested DNA methylation as a feasible biomarker for CAD. Two
19 recent large-scale EWAS surveyed multiple cohorts of various ancestries and
20 collectively reported 85 DNA methylation sites in blood leukocytes to be associated
21 with incident CAD or myocardial infarction (MI) ^{10,11}. Comprehensive studies also
22 report association between DNA methylation and the risk factors of CAD including
23 aging ¹², smoking ¹³, blood lipids ¹⁴, inflammation ¹⁵, hypertension ¹⁶, and diabetes
24 mellitus (DM) ¹⁷. Furthermore, initial EWAS studies identified strong signals that

1 predicted all-cause death of cardiovascular diseases^{18,19}, albeit its biological
2 mechanisms remained to be explored. As such, DNA methylation indicates not only
3 the risk of CAD incidence but also its progression.

4 In this study, we conducted a two-stage multicenter EWAS on prognosis of
5 CAD in 933 patients, profiling blood leukocyte-derived DNA methylation by Illumina
6 MethylationEPIC BeadChip and interrogating its association with patient outcomes in
7 up to 13 years of follow-up. We defined the primary endpoint as all-cause death and
8 the secondary endpoint as major adverse cardiovascular events (MACE), including
9 death, nonfatal myocardial infarction, coronary revascularization, and stroke. From
10 differentially methylated probes/sites (DMPs), we inferred mediating phenotypes,
11 built risk prediction models, assessed the contribution of genetic regulation, and
12 finally, evaluated how the genes impacted by DMPs were expressed during the
13 adverse events. Our results show that DNA methylation of leukocytes from peripheral
14 blood provides robust biomarkers and rich insights into the prognosis of CAD.

15 **Methods**

16 **Cohort assembly and baseline information collection**

17 This study was approved by the Medical Research Ethics Committee of
18 Guangdong Provincial People's Hospital (approval number: GDREC2017071H) and
19 complied with the Declaration of Helsinki. All patients provided written informed
20 consents.

21 We recruited over 5,000 CAD patients from three medical centers in two areas
22 of China for studying the prognosis of CAD, namely Guangdong Provincial People's
23 Hospital, First Affiliated Hospital of Sun Yat-sen University, and Xiangya Hospital of

1 Central South University²⁰. We selected 405 patients recruited between January 2010
2 to December 2017 from Guangdong Provincial People's Hospital to form the
3 discovery cohort, based on a nested case-control study design. 528 patients recruited
4 from 2017 to 2018 from all three medical centers were assembled as the validation
5 cohort. All cohort participants were identified either with a history of coronary artery
6 bypass graft operation or newly diagnosed by coronary angiography and carotid artery
7 ultrasonography to have $\geq 50\%$ obstruction, as assessed by the luminal diameter, in
8 minimally one main coronary artery. The inclusion criteria were: (1) aged over 30
9 years old, (2) no history of renal transplantation or dialysis, (3) no cirrhosis, (4) not
10 pregnant nor breastfeeding, (5) no malignancy, (6) no history of haemodialysis; (7) no
11 history of thyroid problems, not using antithyroid drugs nor thyroid hormone
12 medication in the past week, and (8) completed the follow-up surveys.

13 The cohort participants were admitted to hospitals; after overnight fasting, their blood
14 samples were drawn at 7AM on the second morning. Clinical laboratory tests were
15 performed and detailed clinical surveys, including medical history, family history,
16 smoking status, and medication intake were collected as baseline information.
17 Echocardiography was used to determine the function and structure of the left
18 ventricle (LV). All patients were followed up by telephone every six months by the
19 medical staff team for inquiring about occurrences of all-cause death or major adverse
20 cardiovascular events (MACE), with the latter defined as nonfatal myocardial
21 infarction, coronary revascularization, stroke, and death.

22 **DNA extraction from blood leukocytes**

23 Whole blood was collected in EDTA-K2 anticoagulant tubes and immediately
24 separated into plasma and hemocyte by centrifuging at 1000 g for 10 min at 4°C.

1 Genomic DNA was extracted from hemocyte and transferred to cryopreservation
2 tubes, which were stored at -80°C for subsequent experiments.

3 **Genome-wide DNA methylation profiling and data preprocessing**

4 DNA quality was assessed by ultraviolet spectrophotometer (Thermo Scientific,
5 NanoDrop 2000). Briefly, about 500 ng of DNA was treated with sodium bisulfite for
6 converting unmethylated nucleotide C to U, using the EZ DNA Methylation Kit
7 (Zymo Research). After the conversion, methylation levels of more than 850,000 CpG
8 sites were quantified using the Illumina Infinium MethylationEPIC BeadChip, which
9 was run on an Illumina iScan Systems according to the manufacturer's standard
10 protocol. DNA methylation profiling was serviced by Genenergy Inc. The
11 experimental operator was blind to the group information and randomly assigned the
12 samples to different chips and plates.

13 Raw signal intensities of DNA methylation were stored in .idat files and imported to
14 the R environment using the "ChAMP" package ^{21,22}. Analysis was performed
15 separately for the discovery cohort and the validation cohort. Methylation level of
16 each probe, i.e., beta value, was defined as $\text{Meth}/(\text{Meth} + \text{Unmeth} + 100)$, where
17 Meth was signal intensity of the CpG site in methylated form and Unmeth was that in
18 unmethylated form. Beta values ranged from 0 to 1, with a larger value indicating a
19 higher level of methylation. Probes were excluded if meeting one of the following
20 criteria: (1) with detection P value ≥ 0.01 , (2) with beadcount < 3 in at least 5% of
21 samples, (3) DNA methylation occurring to non-CpG dinucleotides, (4) aligning to
22 multiple locations ²³, (5) located on chromosome X or Y. In total, 733,638 probes in
23 the discovery cohort and 738,366 probes in the validation cohort were retained.

1 The qualified probes were normalized with the BMIQ method ²⁴ to correct for signal
2 bias caused by type-I and type-II probes on the array. Next, we used the method
3 proposed by Houseman *et al.* ²⁵ to estimate the relative proportions of blood cells,
4 including CD8 lymphocytes, CD4 lymphocytes, natural killer cells, B cells,
5 monocytes, and granulocytes. We also leveraged 224 positive control probes to
6 evaluate the impact of technical confounders, which we generally referred as batch
7 effect, on the DNA methylation values. Briefly, we computed the principal
8 components (PCs) of these positive control probes and assessed the association
9 between the first 20 PCs and several technical parameters, including the indices for
10 bisulfite conversion batch, plates, sample wells and chip. Methylation residuals were
11 then obtained via linear regression, with the beta value of each probe as independent
12 variable, and age, sex, smoking status, estimated white-blood-cell proportions, and the
13 top 10 PCs of positive control probes as dependent variables.

14 **Epigenome-wide association analysis**

15 Cox regression-based survival analysis was employed to explore the association
16 between each methylation residual and the trait, i.e., all-cause death or MACE. We
17 performed such EWAS separately for the discovery and the validation cohorts. In
18 each EWAS, we adjusted for age, sex, smoking status, percutaneous coronary
19 intervention (PCI), arrhythmia, heart failure, hypertension, hyperlipidemia, and
20 medication intake including β -receptor blocker (BB), angiotensin converting enzyme
21 inhibitors (ACEI), calcium channel blocker (CCB), proton pump inhibitor (PPI),
22 clopidogrel, and aspirin. A strict epigenome-wide significance threshold by
23 Bonferroni correction was set as $P < 6.83E-08$ and a moderate threshold by Benjamini
24 & Hochberg correction was set as $FDR < 0.05$. The differentially methylated site was
25 considered validated when the association showed a consistent direction of effect in

1 both cohorts, with $FDR < 0.05$ in the discovery cohort and $P < 0.05$ in the validation
2 cohort.

3 **Characterizing genomic locations of DMPs**

4 Genomic locations of DMPs were annotated by Annovar²⁶. Overlap with regulatory
5 elements were computed based on ENCODE Encyclopedia version 5 (ENCODE5)
6 cCRE catalog²⁷, including insulators, promoters, distal enhancers, and proximal
7 enhancers. Enrichment against tissue- and cell type- specific regulatory elements was
8 performed based on histone modification chromatin immunoprecipitation peaks (ChIP)
9 (H3K4me1, H3K4me3, H3K27me3, H3K36me3, H3K9me3, and H3K27ac marks)
10 and regions of 15 chromatin states across 299 cell types and tissues from Roadmap
11 Epigenomics^{28,29} in eFORGE v2.0 (<https://eforge.altiusinstitute.org/>).

12 **Target gene predictions**

13 We predicted the target genes impacted by the DMPs by two methods. For one, we
14 used the annotation file provided by Illumina, which assigned each CpG site to its
15 nearest gene. For the other, we leveraged the activity-by-contact (ABC) model
16 developed by Nasser et al.^{30,31}, which identified active enhancers in a particular cell
17 type and predicted their target genes based on chromatin states and three-dimensional
18 contacts. To identify the ABC enhancers that overlap with DMPs, we adopted the
19 GWAS annotation approach by Zhang K et al.³² by adding ± 2500 bp to the genomic
20 location of DMPs and overlapped them with the ABC enhancers of 131 human cell
21 types. We adopted the original ABC score thresholds, namely ≥ 0.015 for distal
22 element-gene connections and ≥ 0.1 for proximal promoter-gene connections, to
23 define DMP – enhancer – target gene connections.

1 Construction of prognostic models for death and MACE

2 Risk prediction models were constructed based on the discovery cohort and tested in
3 the validation cohort. For building the methylation model of death, 21 DMPs passing
4 the Bonferroni-corrected epigenome-wide significance threshold were pruned by a
5 random forest approach (feature pruning) and those retained were fit by the
6 multivariate Cox regression to derive the final model (weight tuning). For feature
7 pruning, the parameters m_{try} and n_{tree} in the random forest models were tuned using
8 the out of bag error for deriving a minimal overall misclassification rate. m_{try} refers to
9 the number of variables tried at each split and n_{tree} refers to the number of trees to be
10 grown in a forest. The top 10 DMPs with the largest variable importance measure
11 (VIM), which denoted the contribution of each input feature to the model, were
12 retained. For the risk model of MACE, all eight DMPs passing the Bonferroni-
13 corrected epigenome-wide significance threshold were retained. In weight tuning, the
14 retained DMPs were fit by multivariate Cox regression in the R package “survival”.
15 For deriving robust AUC (The area under the receiver operating characteristic (ROC)
16 curve) values, we adopted a process of 80:20 data split and 1,000 times cross
17 validation. The final model was obtained by combining all patients in the discovery
18 cohort.

19 We applied two types of models to the multi-center validation cohort for testing their
20 performances, with one combining CpG sites, sex, and age, and the other combining
21 CpG sites, sex, age, HDLC, Fibrinogen, and LVEF. Prediction risk scores for five-
22 year survival were computed, and Wilcoxon test was used to assess whether the
23 scores between the two groups of patients, i.e., with and without the events of death or
24 MACE, were significantly different. The sensitivity and specificity of these two

1 models were computed using the ConfusionMatrix function from the R package
2 “caret”.

3 Several clinical features were also assessed for their capability in predicting CAD
4 prognostic outcomes, alone or in combination with other clinical features and the
5 selected methylation sites, using the R package “survival”.

6 **Pleiotropic association analysis of DMPs and eQTL**

7 Summary-data based Mendelian Randomization (SMR) analysis³³ and the
8 Heterogeneity in Dependent Instruments (HEIDI) test were employed to identify
9 pleiotropic relationships between the DMPs and gene expression
10 (<https://yanglab.westlake.edu.cn/software/smr/#Download>). GWAS summary
11 statistics for DNA methylation in Asian populations, and therefore information of
12 methylation quantitative trait loci (meQTL), were obtained from Peng et al³⁴. The
13 association strength *beta* theoretically ranges between -1 and 1 for maximally
14 negative to maximally positive associations. The information of *cis*-eQTL were
15 downloaded from eQTLGen (<https://www.eqtlgen.org/>). Allele frequencies were
16 obtained by referring to the East Asians in the 1000 Genomes Project reference panel
17 (phase3, version5)³⁵. 69 DMPs with at least one *cis*-meQTL ($P < 1.0E-08$) were
18 selected for computing causal relationships with *cis*-regulated gene expression. The
19 CpG-gene expression associations with a Bonferroni-corrected P value ($P < 0.05/69 =$
20 $7.25E-04$) were further selected for the HEIDI test ($P_{\text{HEIDI}} > 0.05$) to distinguish
21 pleiotropy from linkage.

1 **Genome-wide association study of death and MACE**

2 We performed genome-wide association studies (GWAS) on death and MACE in
3 1,551 CAD patients recruited from Guangdong Provincial People’s Hospital, Xiangya
4 Hospital of Central South University, and the First Affiliated Hospital of Sun Yat-sen
5 University. These patients were genotyped by Illumina Infinium GSA-24 v1.0 bead
6 chip on 700,078 single-nucleotide genomic positions, which, after genotype
7 imputation against the East Asian populations in the 1000 Genomes Project,
8 generated 3,435,397 high-quality single nucleotide variants (SNVs). Details about
9 cohort enrollment, baseline characteristics, data quality control, and genotype
10 imputation were described previously²⁰. Logistic regression was employed for the
11 GWAS via the PLINK software (version 2.0). The first 10 principal components, sex,
12 age, smoking status, percutaneous coronary intervention (PCI), arrhythmia, heart
13 failure, hypertension, hyperlipidemia, and medication intake including β -receptor
14 blocker (BB), angiotensin converting enzyme inhibitors (ACEI), calcium channel
15 blocker (CCB), proton pump inhibitor (PPI), clopidogrel, and aspirin were included as
16 covariates.

17 **Differential gene expression analysis in MI and stroke**

18 We obtained from Kuppe et al³⁶ single-nucleus RNA sequencing data from 19
19 patients with acute MI and four non-transplanted heart donors as controls. A total of
20 191,795 nuclei from 31 tissue samples, including ten major cardiac cell types, were
21 obtained. We performed differential gene expression analysis between the MI patients
22 and controls, as well as among three tissue zones, namely myogenic, ischemic, and
23 fibrotic zones. We also assessed differences between groups by cell types. Wilcoxon
24 tests implemented in the FindMarkers function of the R package “Seurat” were used

1 ³⁷. Genes passing the Bonferroni-corrected P value of 0.05 were considered
2 differentially expressed. Time-series expression analysis based on Fuzzy C-means
3 clustering was used to demonstrate the relative expression changes of prognostic
4 genes.

5 We obtained bulk RNA sequencing of peripheral blood from patients of MI ³⁸ and
6 patients of ischemic stroke ³⁹. Differential gene expression analysis was performed
7 between patients and controls using the R package “limma” ⁴⁰.

8 **Statistical tests**

9 Baseline demographic and clinical characteristics were presented as mean \pm standard
10 deviation for continuous variables and counts (%) for categorical variables. Cox
11 regression-based survival analysis was employed for assessing association between
12 the features and outcomes. Linear regression was used to explore the relationships
13 between DMPs and six inflammatory markers, four blood lipids and two left
14 ventricular indices. Enrichment analysis of biological pathways and traits in GWAS
15 Catalogue database were carried out by R package “enrichR” and terms with a P
16 value smaller than 0.05 was considered as significant. Unless stated, P values derived
17 from multiple tests were corrected by methods of FDR or Bonferroni correction.
18 Wilcoxon test was used to assess if the difference of continuous variables between
19 two groups were statistically significance. For counts, chi-square tests were used.

20 **Results**

21 **Baseline characteristics of the study population**

22 We adopted a two-stage multicenter design for studying DNA methylation
23 related to CAD prognosis (*Figure 1a*). We assembled over 5,000 CAD patients from a

1 large medical center in China and based on the nested case-control study design,
2 selected 405 patients to form the discovery cohort. As such, a total of 217 deaths and
3 247 MACE events were recorded in up to 13 years of follow-up, while 158 patients
4 experienced no adversity. For the validation cohort, we chose a forward study design
5 and enrolled 528 CAD patients from three medical centers in China. We followed the
6 patients in the validation cohort for about 3 years and observed 25 deaths and 41
7 MACE events.

8 The baseline characteristics of the patients at the time of enrollment were
9 expounded in [Supplementary data online, Table S1](#) and [Figure S1](#). Consistent with
10 epidemiological observations, our CAD patients were mainly men aged over 60 years
11 (73%). We documented their basic demographic information, medical history,
12 biomedical measurements after overnight fasting, and medication intake during the
13 ascertained periods. The measurements were comparable between the two cohorts ($P >$
14 0.05), except for a few, most of which were included as covariates in our statistical
15 tests. Overall, the two cohorts displayed similar patient characteristics. The follow-up
16 times and event times are displayed in [Supplementary data online, Figure S1](#).

17 **EWAS identified differentially methylated CpGs for death and MACE**

18 We extracted DNA from leukocytes in peripheral blood and performed
19 genome-wide methylation profiling via the Illumina Infinium MethylationEPIC
20 BeadChip. After stringent quality control (see [Supplementary data online, Figure S2](#)),
21 we obtained 733,737 and 738,021 high-quality CpG probes from the discovery and
22 the validation cohorts, respectively.

23 Next, we employed Cox regression for EWAS on the high-quality DNA
24 methylation sites against the survival of all-cause death or MACE (see [Supplementary](#)

1 data online, [Figure S3a-S3b](#) and [Table S2-S4](#)). In the discovery cohort, after
2 correcting for sex, age, smoking status, percutaneous coronary intervention, heart
3 failure, hypertension, arrhythmia, hyperlipidemia, and medications, a total of 554
4 differentially methylated CpGs (DMPs) passed the moderate significance threshold
5 (FDR adjusted P , i.e., $Q < 0.05$, [Figure 1b](#)). In the validation cohort, 105 of the 554
6 DMPs were replicated, defined by $P < 0.05$ and consistent direction of effect.
7 Remarkably, 21 DMPs in the discovery cohort remained significant after the
8 epigenome-wide Bonferroni correction ($P < 6.83E-08$), and six of them were
9 replicated in the validation cohort ($P < 0.05$). Similarly, we performed EWAS on the
10 secondary endpoint, MACE. 30 out of the 95 DMPs were validated, among which
11 five passed the Bonferroni-corrected threshold.

12 Overall, 115 DMPs were detected in both cohorts as associated with adverse
13 outcomes of CAD, with most showing increased methylation ([Figure 2a](#),
14 [Supplementary data online, Figure S3 and Table S5](#)). Notably, 60% DMPs were
15 reported in EWAS Catalog⁴¹ and EWAS Atlas⁴² to connect to a variety of traits and
16 diseases, including Crohn's disease and inflammatory bowel disease (25 DMPs),
17 smoking (16 DMPs), aging (8 DMPs), and C-reactive proteins (6 DMPs)¹⁵.
18 Associations with body mass index^{17,43}, atherosclerotic plaque⁴⁴, and death risk⁴⁵
19 were also observed.

20 These 115 DMPs tend to reside in non-coding regions, particularly distal
21 enhancers (see [Supplementary data online, Figure S4a-S4b](#)). Overlap with histone
22 modification ChIP peaks and the 15 chromatin states from Roadmap (**Methods**)
23 revealed that these DMPs were strongly enriched in enhancers specific to blood
24 monocytes, adipocytes, myoepithelial cells, and fibroblasts. Furthermore, the left
25 ventricle and right atrium also appeared on the top of the list ($Q < 0.05$, see

1 [Supplementary data online, Figure S4c-S4d](#)). These results suggest that the DMPs
2 were prone to occur in regions characteristic of heart traits. Note that our DNA
3 methylation were derived from blood leukocytes; however, the DMPs tend to locate
4 in enhancers characteristic of not only blood but also tissues and cells of non-blood
5 origin yet known to play critical roles in CAD. Given that both DNA methylation and
6 RNA transcription in blood were distinct from solid organs ⁴⁶, our observation
7 suggested that methylation in leukocytes indeed carried pathophysiological features
8 and therefore suitable to serve as biomarkers

9 A main function of methylation is to regulate gene expression. By annotating
10 DMPs to the closest genes, we discovered 100 prognostic genes, which were highly
11 enriched for apoptotic pathways, stress response, inflammation response, and vascular
12 processes ([Figure 2b](#)). Furthermore, these genes were enriched for CAD and immune
13 traits as recorded in the GWAS Catalog ([Figure 2c](#)).

14 **Pathways and mediating phenotypes of DMPs**

15 We also interrogated target genes by querying the connections of enhancers to
16 target genes via the Activity-by-Contact (ABC) model, which leveraged chromatin
17 states and three-dimensional contacts (**Methods**). For linking DMPs to enhancers, we
18 adopted a liberal approach in Boix et al. (**Methods**) for connecting GWAS loci to
19 enhancers, by which we extended the DMPs by adding 2500 bp to both flanking
20 regions. Next, we assessed if the 5 Kb regions overlapped with any enhancers, which
21 in turn were connected to genes via the ABC model. In this way, 83 of the 115 DMPs
22 (72.1%) were connected to 806 genes (see [Supplementary data online, Table S6](#)).
23 Consistent with the closest gene annotation, these genes were enriched for
24 inflammation and cellular senescence ($P < 0.05$). Notably, three methylation sites,
25 cg25114611, cg25500196, and cg25563198, were each predicted to interact with > 50

1 genes. These big gene clusters were strongly enriched in stress-induced senescence
2 and inflammation response ($Q < 0.05$) (see [Supplementary data online, Figure S5](#)).
3 Furthermore, all three DMPs were located on super-enhancers active in CAD relevant
4 tissues, such as blood, lymphoid, adipose tissue, heart ventricle, and aorta ⁴⁷.

5 Theoretically, the ABC target genes entail cell type-specific functions, due to
6 enhancers' selective activity in different cells. Among the blood cells, the target genes
7 displayed an enrichment for JAK-STAT and interleukin (IL) pathways (IL-2, IL-7,
8 IL-9, and IL-15) in monocytes and T helper cells, whereas an enrichment for IL-17
9 pathway was found active in more diverse cell types, including monocytes,
10 macrophages, CD4+ T helper cells, and CD19+ B cells (see [Supplementary data
11 online, Table S7](#)). We also extended the analysis to all 131 human cell types and
12 tissues in the ABC model and found these immune pathways active in 24% of the
13 tissues and cell, including those closely related to CAD, e.g., coronary artery, adipose,
14 liver, epithelium, and T cells (see [Supplementary data online, Figure S6](#)), suggesting
15 the adverse outcomes involved robust immune response pathways. There also
16 appeared to be a marginal enrichment of lipid response in T cells and mitochondrial
17 processes in monocytes and dendritic cells ($P < 0.05$).

18 We verified the connection of the DMPs to inflammation and lipids in clinical
19 measurements (see [Figure 2d](#) and [Supplementary data online, Table S8](#)). In both the
20 discovery and the validation cohorts and out of the 115 DMPs, 43 DMPs displayed
21 association with *systemic immune-inflammation index* (SII) as defined by neutrophil
22 count \times platelet count / lymphocyte count ⁴⁸, 37 DMPs associated with a chronic low-
23 grade inflammation index *fibrinogen* ⁴⁹, and 15 DMPs associated with a marker for
24 acute inflammation prothrombotic status *platelet-lymphocyte ratio* (PLR) ⁵⁰. Several
25 other inflammation markers were also investigated; however, the associations were

1 comparably weaker and non-consistent. For lipids, a strong connection of 29 DMPs to
2 high-density lipoprotein cholesterol (HDLC) was observed. Interestingly, few
3 connections were found for low-density lipoprotein cholesterol (LDLC), total
4 cholesterol, or triglycerides (TG). We also examined left ventricular ejection fraction
5 (LVEF) and left ventricular mass index (LVMI), as low LVEF and high LVMI were
6 indicative of LV remodeling and therefore increased risk of death or MACE in CAD.
7 25 and 15 DMPs were respectively associated with LVEF and LVMI, and
8 reassuringly, most of these associations displayed opposite directions for both traits.

9 **Prognostic models for death and MACE in CAD**

10 Risk prediction models help to identify CAD patients with greater risk of developing
11 adverse outcomes. We constructed models based on DMPs that displayed epigenome-
12 wide significance of association in the discovery cohort (n=405). We leveraged a
13 random forest approach to select DMPs that contributed the most to the classification
14 accuracy, and derived precise weights by the multivariable Cox regression-based
15 survival analysis (**Methods**, see [Supplementary data online, Figure S7](#)). As such, we
16 constructed prognostic models for death with 10 CpGs and for MACE with 8 CpGs,
17 which we termed the CG prognostic models ([Table 1](#)). This model for death achieved
18 an area under the curve (AUC) of 0.70 in the discovery cohort, close to or better than
19 using traditional risk factors including sex (AUC = 0.52), chronological age (AUC =
20 0.72) and their combination (AUC = 0.72). Combining sex, chronological age, and
21 the 10 CpGs improved the prediction power (AUC = 0.80) ([Figure 3a](#)). We also built
22 prognostic models based on the mediating phenotypes. Although not all the clinical
23 features were equally powerful in predicting the adverse outcomes (see
24 [Supplementary data online, Figure S8a-S8b](#)), we found that the ensemble model
25 combining the 10 DMPs, sex, age, fibrinogen, HDLC, and LVEF achieved an AUC of

1 0.83 (*Figure 3b*). When applying the models to the validation cohort (n=528), which
2 was assembled from three medical centers and independent of the discovery cohort in
3 which the models were built, we observed about 11% drop in both sensitivity and
4 specificity. Still, the ensemble model could well stratify CAD patients of different
5 risks of death in five years ($P < 0.0001$ (see *Figure 3c-3d* and [Supplementary data](#)
6 [online, Table S9](#)).

7 Similarly, the CG prognostic model for MACE performed the best when
8 combining the 8 CpG sites with sex, age, fibrinogen, HDLC, and LVEF (AUC = 0.77)
9 (*Figure 3e-3f*) and achieved a good patient stratification of five-year risk in the
10 validation cohort (see *Figure 3g-3h* and [Supplementary data online, Table S10](#)).

11 Age is known as a strong risk factor for CVD. Observing the chronological
12 age predicted closely to the CG models, we explored the performance of several DNA
13 methylation clocks, including GrimAge⁵¹, PhenoAge⁵², Hannum Clock⁵³, and
14 Horvath Clock⁵⁴. DNA methylation clocks have been shown to better represent one's
15 aging status. Indeed, most clock models achieved better prediction than the
16 chronological age models for death and MACE (see [Supplementary data online,](#)
17 [Figure S8c-S8d](#)) and performed equally well or even better than the CG prognostic
18 models. As these clocks comprised dozens to hundreds of CpG sites, one to two
19 orders of magnitude more than the maximal 10 CpG sites in our models, our CG
20 prognostic models are more succinct and specific.

21 Note that among our CAD patients with adverse outcomes, a vast majority had
22 the events occurred within the first 5 years of the ascertained period, therefore our
23 models mostly captured the risk of adversity in a relative near term.

1 **Contribution of genetic regulation to CAD prognosis**

2 Methylation can be regulated genetically by quantitative trait loci (meQTL), thus
3 providing a tool for investigating how genetics influences CAD prognosis. We
4 queried the 115 DMPs against a meQTL dataset derived from 3,523 East Asians
5 (**Methods**). Compared with all CpG sites on the Infinium methylationEPIC beadchip,
6 the DMPs were enriched for both *cis*-meQTL, defined as within 1 Mb flanking
7 regions (Hypergeometric $P = 2.12E-08$), and *trans*-meQTL, defined as >5 Mb or
8 SNP-CpG pair located on different chromosomes (Hypergeometric $P = 3.55E-26$).
9 Briefly, 70 DMPs were subjected to the regulation of 14,374 unique *cis*-meQTLs and
10 52 DMPs to 1,612 unique *trans*-meQTLs, with weak associations in most cases.
11 Compared with *trans*, DMPs in *cis* relations tended to have more meQTLs.
12 Altogether 92 of the 115 DMPs (80%) were regulated genetically ([Figure 4a-4d](#)).

13 We next assessed how these meQTLs performed in genetic association tests.
14 We genotyped 1,551 CAD patients using Illumina GSA array, who were recruited
15 from the three medical centers in China and partially overlap with our methylation
16 cohort. We derived 3,448,646 high-quality single nucleotide variants (SNVs) after
17 imputation and quality control, and performed genome-wide association studies
18 (GWAS) with death or MACE via logistic regression (**Methods**). Among the 14,374
19 *cis*-meQTLs paired with the DMPs, 8,362 were genotyped and therein fewer than 5%
20 SNPs displayed nominal association ($P < 0.05$) (see [Supplementary data online](#),
21 [Figure S9a-S9b](#)). Such weak associations would be missed by GWAS, whereas our
22 EWAS analysis recovered them. The fact that 10s to 1000s of meQTLs regulating one
23 single DMP, each with a weak strength, suggested that the trickling of little genetic
24 signals had mounted to significant epigenetic outcomes, which resembled the
25 polygenic model in complex traits. As such, the most important DMPs could be

1 regulated by many meQTLs. Indeed, we observed that the DMP having most cis-
2 meQTLs was cg20015729, with a remarkable 1,842 *cis*-meQTLs spanning 678.84 Kb,
3 and closest to the gene Ubiquitin conjugating enzyme E2 E2 (*UBE2E2*). Similarly,
4 cg16500036 was associated with over 1,000 *cis*-meQTLs, and closest to the gene
5 Activator of transcription and developmental regulator (*AUTS2*). DMPs having most
6 trans-meQTLs were cg25563198 and cg25114611, both closest to the gene FKBP
7 Prolyl Isomerase 5 (*FKBP5*) ([Figure 4e-4f](#))

8 Pleiotropy at the nucleotide level was frequently observed in genetic studies⁵⁵.
9 We were interested in learning the SNVs that simultaneously regulated DMP
10 methylation and gene expression, i.e., SNV with a dual role of meQTL and eQTL.
11 Therefore, we leveraged the pleiotropic association model in SMR (**Methods**) to
12 integrate meQTLs from Peng et al and eQTLs from eQTLGen. Briefly, 6,796 *cis*-
13 meQTLs connected to 56 DMPs were identified as or closely located to eQTLs (P
14 $<1.0E-6$), which regulated the expression of 242 genes. To further distinguish
15 pleiotropy from linkage, we performed a HEIDI test against the null hypothesis that
16 the SMR association was due to pleiotropy. As such, we revealed that 1,785 *cis*-
17 meQTLs were indeed eQTLs ($P_{\text{HEIDI}} >0.05$), ruling out the possibility that the
18 pleiotropic associations were caused by genotypes in linkage disequilibrium (LD).
19 These pleiotropic meQTLs/eQTLs were linked to 29 DMPs and 71 genes, forming 80
20 CpG-gene pairs (see [Supplementary data online, Table S11](#)).

21 It is worth noting the SNV rs10235487 (see [Supplementary data online, Figure](#)
22 [S9c](#)). As a *cis*-meQTL, it regulated cg16500036, a common feature in the CG
23 prognostic models of death and MACE. This CpG site was located on an enhancer
24 predicted by the ABC model to interact with *AUTS2*, therefore its increased
25 methylation level would theoretically decrease the expression of *AUTS2*. Meanwhile,

1 as an eQTL, this SNV negatively correlated with the expression of *AUTS2* ($P_{SMR} =$
2 4.50E-09, $P_{HEIDI} > 0.05$). Collectively, these pieces of evidence pointed to a strong
3 connection of SNV rs10235487 to increased methylation of cg16500036 on an
4 enhancer of *AUTS2*, which subsequently results in decreased expression of *AUTS2*,
5 leading to poor CAD outcomes.

6 **Prognostic genes in acute MI and stroke**

7 Our convergent findings presented above indicated that changes of methylation
8 perturbed expression of the key genes involved in regulating CAD progression (see
9 [Supplementary data online, Table S5](#)). To further investigate their effects, we
10 examined expression of the 100 prognostic genes in patients of MI or stroke, two
11 major adverse outcomes of CAD.

12 Bulk RNA sequencing of peripheral blood between MI patients and controls
13 were obtained. In addition, a recent study reported single-nuclei profiling of gene
14 expression and chromatin accessibility with spatial and time resolution, in which three
15 zones of the MI lesion tissues were interrogated. We reanalyzed these two datasets
16 and found that our prognostic genes displayed relatively small expression changes in
17 MI, whether in tissue or blood, as most expression changes were within 30% ([Fig. 5a-](#)
18 [5b](#)). This likely reflects the nature of these genes being more sentinel rather than
19 violent players during the actual occurrence of the adverse events. A closer look
20 showed that 27 of the 100 prognostic genes changed their expression by 30% or more
21 in the 3 physiological zones: *myogenic*, representing non-ischemic or normal tissues;
22 *ischemic*, the lesion site; and *fibrotic*, representing lesion sites with advanced disease
23 progression (see [Supplementary data online, Figure S10a](#) and [Supplementary data](#)
24 [online, Table S12](#)).

1 We also examined expression of the prognostic genes in different cell types
2 (*Figure 5c* and *Supplementary data online, Figure S10b*). Notably, ***FKBP5***, the gene
3 closest to the DMPs cg03546163 (a common marker in CG prognostic models for
4 death and MACE), cg25114611 and cg25563198 (having the largest number of trans-
5 meQTL and ABC target genes), elevated its expression in MI by four folds in nearly
6 all cell types, and its expression levels were ever higher along the disease progression.
7 Conversely, ***AUTS2*** nearest to cg16500036 (a common marker in the CG prognostic
8 models for death and MACE, and serving a dual role as meQTL and eQTL),
9 decreased its expression by 52% and 26% in immature innate lymphoid cells and
10 smooth muscle myoblast cell in MI, and decreased gradually in immature innate
11 lymphoid cells along the disease progression. ***UBE2E2*** nearest to cg20015729 (a
12 marker in the CG prognostic model for death), displayed highest expression in
13 adipocytes of the epicardial fat from the left ventricle; furthermore, we observed its
14 expression was decreased by 20% in immature innate lymphoid cells and native cells
15 of MI, and along the disease progression, it first decreased then elevated in several
16 cell types, including fibroblasts of the cardiac tissue, immature innate lymphoid, and
17 native cells, suggesting this gene may participate in cardiac repair after acute
18 myocardial injury.

19 As for stroke, in a dataset where gene expression in the peripheral blood from
20 patients of acute ischemic stroke was studied, we found 14 prognostic genes
21 displaying differential expression as compared to the healthy controls ($Q < 0.05$)
22 (*Figure 5d*). Therein, ***FKBP5*** was among those having the greatest expression
23 changes.

1 Therefore, there appears to be coherent evidence for the strongest signals, such
2 as those linked to FKBP5, AUTS2, and UBE2E2. We summarized them and
3 constructed the DMP – gene regulation models (*Figure 5e*).

4 **Discussion**

5 In this study, we presented comprehensive EWAS on more than 733,000 DNA
6 methylation sites distributed genome-wide against the survival time of adverse
7 outcomes in 933 CAD patients. We identified 115 CpG sites whose methylation
8 patterns were characteristic of patients with future adverse events. From a spectrum of
9 analyses on these DMPs, we obtained three main observations.

10 First, we learned that 72% of the DMPs were located on enhancers and
11 associated with genes involved in stress response, senescence, inflammation, and
12 vessel tube regulation. An additional investigation leveraging clinical measurements
13 revealed finer sub-categories, from which strong associations of the DMPs were
14 observed with: (1) three inflammation indices, namely fibrinogen, SII, and PLR, all
15 connected to platelets; (2) heart functions, namely LVEF and LVMI; and (3) HDLC.
16 As platelets and cholesterol were essential components of thrombosis, our results
17 suggested that early thrombo-inflammation and heart contraction functions mediated
18 the adverse outcomes in CAD. Indeed, prognostic models based on fibrinogen, LVEF,
19 or LVMI, each achieved AUC 0.65 or above. Interestingly, HDL is known in reverse
20 cholesterol transport, interacts with platelets and exerts an antithrombotic function by
21 suppressing the coagulation cascade and stimulation of clot fibrinolysis⁵⁶. In our
22 analyses, DMPs were associated strongly with HDLC but little with other lipid
23 categories; furthermore, HDLC displayed stronger power than other lipid categories
24 in predicting adverse outcomes. As such, our study suggested that the ability to

1 remove cholesterol, rather than its accumulation, was more relevant to the adverse
2 outcomes. A recent study discovered that LDLC, compared to the inflammation index
3 C-reactive protein, was less effective in predicting future cardiovascular events and
4 death⁵⁷. Our study suggests that HDLC, not LDLC, may be a more relevant predictor.
5 Therefore, our methylation study of CAD adverse outcomes may inspire new research
6 for clinical translation.

7 Second, we observed a significant genetic component in the regulation of
8 CAD adverse outcomes. Strikingly, 80% of the DMPs could be mapped to known
9 meQTLs. In addition, important prognostic genes, which repetitively appeared as
10 most significant in various analyses, own the largest number of meQTLs for their
11 DMPs. For example, *UBE2E2* and *AUTS2*, each had a DMP (cg20015729 and
12 cg16500036) associated with more than 1,000 *cis*-meQTLs. *FKBP5* had two DMPs,
13 with one (cg25563198) associated with 267 *cis*-meQTLs and 273 *trans*-meQTLs, and
14 the other (cg25114611) associated with 270 *trans*-meQTLs. Interestingly, our analysis
15 showed that each of these meQTLs conferred a very weak GWAS signal; however,
16 when combined, they collectively regulated the important DMPs. Indeed, the most
17 important DMPs were regulated under significantly more meQTLs. Such genetic
18 regulation of DNA methylation resembles the polygenic model in GWAS studies of
19 many complex traits⁵⁸. Furthermore, we discovered that pleiotropic effects were
20 general. In total, 55% of the meQTLs (n=6,796) that regulated the adverse outcome
21 DMPs were in LD with known eQTLs. In fact, 15% *cis*-meQTLs (n=1,785) were
22 themselves eQTLs, as assessed by the SMR and HEIDI tests. These observations
23 suggested that a coupled genetic regulation of methylation and gene expression could
24 be a robust mechanism.

1 Third, most prognostic genes displayed subtle changes during the actual
2 adverse outcomes. This can be attributed to the nature of our study, that the
3 biomarkers we searched mainly served as early alarms. These biomarkers represented
4 preceding events months to years before the adverse outcomes actually occurred.
5 That said, however, several prognostic genes did display drastic expression changes
6 during the adverse events and occurred repetitively as the most significant findings
7 along various analyses. These genes include *FKBP5*, *AUTS2* and *UBE2E2*. *FKBP5* is
8 an immunophilin protein that binds to immunosuppressive drugs. In our analysis,
9 *FKBP5* had numerous DMPs associated with inflammation markers and heart
10 functions. One of them, cg25114611, has been reported in acute MI ⁵⁹, death risk ⁴⁵,
11 inflammatory bowel disease ⁶⁰, Crohn's disease ⁶¹, maternal BMI ⁶² and diabetes
12 mellitus ⁶³. *FKBP5* expression was reported to be significantly altered in dilated
13 cardiomyopathy after heart transplantation and suggested to serve as a prognostic
14 marker ⁶⁴. In our analysis of MI and stroke, *FKBP5* appeared as a most highly
15 regulated gene and involved in pathways of cellular senescence. Its elevation in MI
16 tissue was most drastic in innate immune lymphoid cells and adipocytes of the
17 epicardial fat of the left ventricle. These results align with the recent finding that
18 DNA demethylation led to increased expression of *FKBP5*, which in turn promoted
19 NF- κ B signaling in immune cells, resulting in a proinflammatory response and
20 increased cardiovascular risk ⁶⁵. Genetic variation in *AUTS2* was reported in blood
21 pressure ⁶⁶, body mass index ⁶⁶, type 2 diabetes ⁶⁷, and mild heart defects ⁶⁸. *UBE2E2*
22 was associated with type 2 diabetes ⁶⁹, RR interval in electrocardiogram ⁷⁰, and fat
23 distribution ⁷¹. Functional analysis indicated that loss of function of *UBE2E2* in
24 mouse primary adipose progenitor cells impaired adipocyte differentiation ⁷¹. Such

1 ample connections to cardiometabolic diseases strongly support the notion that these
2 prognostic genes played critical roles in the CAD progression.

3 Leveraging the DNA methylation markers and the biological insights from our
4 analyses, we constructed succinct prognostic models for predicting death and MACE
5 in CAD patients. To facilitate clinical application, we purposefully selected
6 maximally 10 CpG sites; furthermore, we incorporated simple demographic
7 information such as chronological age and sex, and a few clinical features that were
8 relatively easy to obtain, including fibrinogen, LVEF, and HDLC. Our ensemble
9 models achieved AUC of 0.83 for predicting death and 0.77 for predicting MACE.
10 Furthermore, they achieved robust performance in CAD patients independently
11 assembled from three medical centers, proving their potential for clinical translation.

12 There are several limitations in our study. First, most of the adverse events
13 occurred to our CAD patients were within the first 5 years of ascertainment, therefore
14 our study captured short-term to intermediate signals. Given a longer interrogation
15 timespan, or a study with adverse events occurred in a longer time span, the DMPs
16 predicting longer-term adverse events could appear. Second, the number of patients
17 experienced adverse outcomes in the validation cohort was relatively small, i.e., there
18 are 248 in the discovery cohort and only 41 in the validation cohort. This limited the
19 power of replication. Indeed, 70-80% DMPs from the discovery cohort could not be
20 replicated in the validation cohort, and therefore removed from the bioinformatic
21 analysis and model construction. Third, although 850K EPIC array could assess CpG
22 methylation genome-wide, many CpG sites were not probed and therefore leaves a
23 large room for future discovery of prognostic markers.

1 To conclude, our study displays the value of leveraging DNA methylation of
2 peripheral blood in predicting future adverse events in CAD patients. Further studies
3 are warranted to investigate the roles of the methylation sites, genes, pathways, and
4 mediating phenotypes implicated in our study for a mechanistic understanding of the
5 CAD adverse outcomes.

6 **Acknowledgements**

7 We thank the CAD patients for their willingness to consent and participate in this
8 scientific study. We thank medical staffs of the Guangdong Provincial People's
9 Hospital, particularly the nurses, for their diligent and careful follow-up of the
10 patients. We thank Dr. Sijia Wang for providing early access to the methylation QTL
11 datasets for our research. We also thank members of the Guangdong Provincial Key
12 Laboratory of Coronary Heart Disease Prevention and members of the Laboratory of
13 Intelligent Computing in Biomedicine in the Greater Bay Area Institute of Precision
14 Medicine (Guangzhou) for insightful discussions and suggestions.

15 **Funding**

16 This study was supported by the National Natural Science Foundation of China (No.
17 82274016, 32270626, 81872934), the Key-Area Research and Development Program
18 of Guangdong Province, China (No. 2019B020229003), Greater Bay Area Research
19 Institute of Precision Medicine (Guangzhou) Research Grants (I0005, R2001),
20 National key research and development program (No. 2017YFC0909301) and the
21 Science and Technology Program of Guangzhou (No. 2023B03J1251).

22 **Conflict of interest**

23 All authors declare no competing interests.

1 **Data availability**

2 Gene regulatory elements were obtained from ENCODE5 catalog
3 (<https://www.encodeproject.org/>). Enhancer-gene predictions by the ABC models in
4 131 biosamples were obtained from the ENGREITZ LAB
5 (<https://www.engreitzlab.org/abc/>). CpG sites associated with diseases and traits were
6 download from EWAS Catalog (<http://www.ewascatalog.org/>) and EWAS Atlas
7 (<https://ngdc.cncb.ac.cn/ewas/atlas>). meQTL summary statistic data were obtained
8 from Pan-mQTL (<https://www.biosino.org/panmqt/home>). eQTL summary statistics
9 were derived from eQTLGen (<https://www.eqtlgen.org/cis-eqtls.html>). Single-nuclei
10 RNA sequencing data of MI patients were available from Zenodo data archive
11 (<https://zenodo.org/record/6578047>). Blood derived bulk RNA sequencing data were
12 obtained from GEO database ([GSE61144](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE61144) and [GSE16561](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE16561)). Calculation of DNA
13 methylation age of GrimAge, Hannum clock, Hovath clock, and PhenoAge were
14 performed by DNA Methylation Age Calculator
15 (<https://dnamage.genetics.ucla.edu/home>).

16 **Author Contributions**

17 SZ and CP designed the study. SZ supervised and coordinated the overall study and
18 CP supervised the data analysis. SZ, XF, BZ, XC and CL assembled the study cohort.
19 MQ, QW, XT, QZ, XW, XC and CL consented the patients and supervised the patient
20 follow-up. MQ, QW, XT and HL collected samples and prepared them for DNA
21 methylation array. MQ and CP performed bioinformatic and statistical analyses and
22 generated the figures and tables. CP and MQ drafted the manuscript. All authors
23 contributed to result interpretation and discussions. CP, MQ and SZ critically
24 reviewed the manuscript.

1 Ethical approval

2 This study was approved by the Medical Research Ethics Committee of Guangdong
3 Provincial People's Hospital (approval number: GDREC2017071H) and complied
4 with the Declaration of Helsinki. All patients provided written informed consents.

5 References

- 6 1. Li S, Liu Z, Joseph P, Hu B, Yin L, Tse LA, *et al.* Modifiable risk factors
7 associated with cardiovascular disease and mortality in China: a PURE substudy. *Eur*
8 *Heart J* 2022;**43**:2852-2863. doi: 10.1093/eurheartj/ehac268
- 9 2. Tsao CW, Aday AW, Almarazgoq ZI, Anderson CAM, Arora P, Avery CL, *et*
10 *al.* Heart Disease and Stroke Statistics-2023 Update: A Report From the American
11 Heart Association. *Circulation* 2023;**147**:e93-e621. doi:
12 10.1161/CIR.0000000000001123
- 13 3. Zhou M, Wang H, Zeng X, Yin P, Zhu J, Chen W, *et al.* Mortality, morbidity,
14 and risk factors in China and its provinces, 1990-2017: a systematic analysis for the
15 Global Burden of Disease Study 2017. *Lancet* 2019;**394**:1145-1158. doi:
16 10.1016/S0140-6736(19)30427-1
- 17 4. Cole JH, Miller JI, 3rd, Sperling LS, Weintraub WS. Long-term follow-up of
18 coronary artery disease presenting in young adults. *J Am Coll Cardiol* 2003;**41**:521-
19 528. doi: 10.1016/s0735-1097(02)02862-0
- 20 5. Proudfit WJ, Brusckie AV, MacMillan JP, Williams GW, Sones FM, Jr.
21 Fifteen year survival study of patients with obstructive coronary artery disease.
22 *Circulation* 1983;**68**:986-997. doi: 10.1161/01.cir.68.5.986
- 23 6. Brown JC, Gerhardt TE, Kwon E. Risk Factors For Coronary Artery Disease.
24 In: StatPearls. Treasure Island (FL); 2023.
- 25 7. Ananth CV, Rutherford C, Rosenfeld EB, Brandt JS, Graham H, Kostis WJ, *et*
26 *al.* Epidemiologic trends and risk factors associated with the decline in mortality from
27 coronary heart disease in the United States, 1990-2019. *Am Heart J* 2023;**263**:46-55.
28 doi: 10.1016/j.ahj.2023.05.006
- 29 8. Drobni ZD, Kolossvary M, Karady J, Jermendy AL, Tarnoki AD, Tarnoki DL,
30 *et al.* Heritability of Coronary Artery Disease: Insights From a Classical Twin Study.
31 *Circ Cardiovasc Imaging* 2022;**15**:e013348. doi:
32 10.1161/CIRCIMAGING.121.013348
- 33 9. Zdravkovic S, Wienke A, Pedersen NL, Marenberg ME, Yashin AI, De Faire
34 U. Heritability of death from coronary heart disease: a 36-year follow-up of 20 966
35 Swedish twins. *J Intern Med* 2002;**252**:247-254. doi: 10.1046/j.1365-
36 2796.2002.01029.x
- 37 10. Agha G, Mendelson MM, Ward-Caviness CK, Joehanes R, Huan T, Gondalia
38 R, *et al.* Blood Leukocyte DNA Methylation Predicts Risk of Future Myocardial
39 Infarction and Coronary Heart Disease. *Circulation* 2019;**140**:645-657. doi:
40 10.1161/CIRCULATIONAHA.118.039357
- 41 11. Navas-Acien A, Domingo-Relloso A, Subedi P, Riffo-Campos AL, Xia R,
42 Gomez L, *et al.* Blood DNA Methylation and Incident Coronary Heart Disease:

- 1 Evidence From the Strong Heart Study. *JAMA Cardiol* 2021;**6**:1237-1246. doi:
2 10.1001/jamacardio.2021.2704
- 3 12. Zhang W, Song M, Qu J, Liu GH. Epigenetic Modifications in Cardiovascular
4 Aging and Diseases. *Circ Res* 2018;**123**:773-786. doi:
5 10.1161/CIRCRESAHA.118.312497
- 6 13. Joehanes R, Just AC, Marioni RE, Pilling LC, Reynolds LM, Mandaviya PR,
7 *et al.* Epigenetic Signatures of Cigarette Smoking. *Circ Cardiovasc Genet*
8 2016;**9**:436-447. doi: 10.1161/CIRCGENETICS.116.001506
- 9 14. Dekkers KF, van Ijzerman M, Sliker RC, Moed MH, Bonder MJ, van Galen M,
10 *et al.* Blood lipids influence DNA methylation in circulating cells. *Genome Biol*
11 2016;**17**:138. doi: 10.1186/s13059-016-1000-6
- 12 15. Ligthart S, Marzi C, Aslibekyan S, Mendelson MM, Conneely KN, Tanaka T,
13 *et al.* DNA methylation signatures of chronic low-grade inflammation are associated
14 with complex diseases. *Genome Biol* 2016;**17**:255. doi: 10.1186/s13059-016-1119-5
- 15 16. Richard MA, Huan T, Ligthart S, Gondalia R, Jhun MA, Brody JA, *et al.*
16 DNA Methylation Analysis Identifies Loci for Blood Pressure Regulation. *Am J Hum*
17 *Genet* 2017;**101**:888-902. doi: 10.1016/j.ajhg.2017.09.028
- 18 17. Wahl S, Drong A, Lehne B, Loh M, Scott WR, Kunze S, *et al.* Epigenome-
19 wide association study of body mass index, and the adverse outcomes of adiposity.
20 *Nature* 2017;**541**:81-86. doi: 10.1038/nature20784
- 21 18. Abdulrahim JW, Kwee LC, Grass E, Siegler IC, Williams R, Karra R, *et al.*
22 Epigenome-Wide Association Study for All-Cause Mortality in a Cardiovascular
23 Cohort Identifies Differential Methylation in Castor Zinc Finger 1 (CASZ1). *J Am*
24 *Heart Assoc* 2019;**8**:e013228. doi: 10.1161/JAHA.119.013228
- 25 19. Zhang Y, Wilson R, Heiss J, Breitling LP, Saum KU, Schottker B, *et al.* DNA
26 methylation signatures in peripheral blood strongly predict all-cause mortality. *Nat*
27 *Commun* 2017;**8**:14617. doi: 10.1038/ncomms14617
- 28 20. Wang Z, Zhu Q, Liu Y, Chen S, Zhang Y, Ma Q, *et al.* Genome-wide
29 association study of metabolites in patients with coronary artery disease identified
30 novel metabolite quantitative trait loci. *Clin Transl Med* 2021;**11**:e290. doi:
31 10.1002/ctm2.290
- 32 21. Morris TJ, Butcher LM, Feber A, Teschendorff AE, Chakravarthy AR,
33 Wojdacz TK, *et al.* ChAMP: 450k Chip Analysis Methylation Pipeline.
34 *Bioinformatics* 2014;**30**:428-430. doi: 10.1093/bioinformatics/btt684
- 35 22. Tian Y, Morris TJ, Webster AP, Yang Z, Beck S, Feber A, *et al.* ChAMP:
36 updated methylation analysis pipeline for Illumina BeadChips. *Bioinformatics*
37 2017;**33**:3982-3984. doi: 10.1093/bioinformatics/btx513
- 38 23. Nordlund J, Backlin CL, Wahlberg P, Busche S, Berglund EC, Eloranta ML,
39 *et al.* Genome-wide signatures of differential DNA methylation in pediatric acute
40 lymphoblastic leukemia. *Genome Biol* 2013;**14**:r105. doi: 10.1186/gb-2013-14-9-r105
- 41 24. Teschendorff AE, Marabita F, Lechner M, Bartlett T, Tegner J, Gomez-
42 Cabrero D, *et al.* A beta-mixture quantile normalization method for correcting probe
43 design bias in Illumina Infinium 450 k DNA methylation data. *Bioinformatics*
44 2013;**29**:189-196. doi: 10.1093/bioinformatics/bts680
- 45 25. Houseman EA, Accomando WP, Koestler DC, Christensen BC, Marsit CJ,
46 Nelson HH, *et al.* DNA methylation arrays as surrogate measures of cell mixture
47 distribution. *BMC Bioinformatics* 2012;**13**:86. doi: 10.1186/1471-2105-13-86
- 48 26. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic
49 variants from high-throughput sequencing data. *Nucleic Acids Res* 2010;**38**:e164. doi:
50 10.1093/nar/gkq603

- 1 27. Consortium EP, Moore JE, Purcaro MJ, Pratt HE, Epstein CB, Shores N, *et al.* Expanded encyclopaedias of DNA elements in the human and mouse genomes. *Nature* 2020;**583**:699-710. doi: 10.1038/s41586-020-2493-4
- 2 *al.* Expanded encyclopaedias of DNA elements in the human and mouse genomes.
- 3 *Nature* 2020;**583**:699-710. doi: 10.1038/s41586-020-2493-4
- 4 28. Bernstein BE, Stamatoyannopoulos JA, Costello JF, Ren B, Milosavljevic A, Meissner A, *et al.* The NIH Roadmap Epigenomics Mapping Consortium. *Nat Biotechnol* 2010;**28**:1045-1048. doi: 10.1038/nbt1010-1045
- 5 Meissner A, *et al.* The NIH Roadmap Epigenomics Mapping Consortium. *Nat*
- 6 *Biotechnol* 2010;**28**:1045-1048. doi: 10.1038/nbt1010-1045
- 7 29. Roadmap Epigenomics C, Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, *et al.* Integrative analysis of 111 reference human epigenomes. *Nature*
- 8 A, *et al.* Integrative analysis of 111 reference human epigenomes. *Nature*
- 9 2015;**518**:317-330. doi: 10.1038/nature14248
- 10 30. Nasser J, Bergman DT, Fulco CP, Guckelberger P, Doughty BR, Patwardhan TA, *et al.* Genome-wide enhancer maps link risk variants to disease genes. *Nature*
- 11 TA, *et al.* Genome-wide enhancer maps link risk variants to disease genes. *Nature*
- 12 2021;**593**:238-243. doi: 10.1038/s41586-021-03446-x
- 13 31. Fulco CP, Nasser J, Jones TR, Munson G, Bergman DT, Subramanian V, *et al.* Activity-by-contact model of enhancer-promoter regulation from thousands of CRISPR perturbations. *Nat Genet* 2019;**51**:1664-1669. doi: 10.1038/s41588-019-0538-0
- 14 Activity-by-contact model of enhancer-promoter regulation from thousands of
- 15 CRISPR perturbations. *Nat Genet* 2019;**51**:1664-1669. doi: 10.1038/s41588-019-
- 16 0538-0
- 17 32. Boix CA, James BT, Park YP, Meuleman W, Kellis M. Regulatory genomic circuitry of human disease loci by integrative epigenomics. *Nature* 2021;**590**:300-307. doi: 10.1038/s41586-020-03145-z
- 18 circuitry of human disease loci by integrative epigenomics. *Nature* 2021;**590**:300-307.
- 19 doi: 10.1038/s41586-020-03145-z
- 20 33. Zhu Z, Zhang F, Hu H, Bakshi A, Robinson MR, Powell JE, *et al.* Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat Genet* 2016;**48**:481-487. doi: 10.1038/ng.3538
- 21 Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets.
- 22 *Nat Genet* 2016;**48**:481-487. doi: 10.1038/ng.3538
- 23 34. Peng Q, Liu X, Li W, Jing H, Li J, Gao X, *et al.* Analysis of blood methylation quantitative trait loci in East Asians identifies ancestry-specific effects associated with complex trait variation. *Nat Genet* 2023. doi:
- 24 methylation quantitative trait loci in East Asians identifies ancestry-specific effects
- 25 associated with complex trait variation. *Nat Genet* 2023. doi:
- 26 35. Genomes Project C, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, *et al.* A global reference for human genetic variation. *Nature* 2015;**526**:68-74. doi: 10.1038/nature15393
- 27 A global reference for human genetic variation. *Nature* 2015;**526**:68-74.
- 28 doi: 10.1038/nature15393
- 29 36. Kuppe C, Ramirez Flores RO, Li Z, Hayat S, Levinson RT, Liao X, *et al.* Spatial multi-omic map of human myocardial infarction. *Nature* 2022;**608**:766-777. doi: 10.1038/s41586-022-05060-x
- 30 Spatial multi-omic map of human myocardial infarction. *Nature* 2022;**608**:766-777.
- 31 doi: 10.1038/s41586-022-05060-x
- 32 37. Hao Y, Hao S, Andersen-Nissen E, Mauck WM, 3rd, Zheng S, Butler A, *et al.* Integrated analysis of multimodal single-cell data. *Cell* 2021;**184**:3573-3587 e3529. doi: 10.1016/j.cell.2021.04.048
- 33 Integrated analysis of multimodal single-cell data. *Cell* 2021;**184**:3573-3587 e3529.
- 34 doi: 10.1016/j.cell.2021.04.048
- 35 38. Park HJ, Noh JH, Eun JW, Koh YS, Seo SM, Park WS, *et al.* Assessment and diagnostic relevance of novel serum biomarkers for early decision of ST-elevation myocardial infarction. *Oncotarget* 2015;**6**:12970-12983. doi: 10.18632/oncotarget.4001
- 36 Assessment and diagnostic relevance of novel serum biomarkers for early decision of ST-elevation
- 37 myocardial infarction. *Oncotarget* 2015;**6**:12970-12983. doi:
- 38 10.18632/oncotarget.4001
- 39 39. Barr TL, Conley Y, Ding J, Dillman A, Warach S, Singleton A, *et al.* Genomic biomarkers and cellular pathways of ischemic stroke by RNA gene expression profiling. *Neurology* 2010;**75**:1009-1014. doi: 10.1212/WNL.0b013e3181f2b37f
- 40 Genomic biomarkers and cellular pathways of ischemic stroke by RNA gene
- 41 expression profiling. *Neurology* 2010;**75**:1009-1014. doi:
- 42 10.1212/WNL.0b013e3181f2b37f
- 43 40. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* 2015;**43**:e47. doi: 10.1093/nar/gkv007
- 44 differential expression analyses for RNA-sequencing and microarray studies. *Nucleic*
- 45 *Acids Res* 2015;**43**:e47. doi: 10.1093/nar/gkv007
- 46 41. Battram T, Yousefi P, Crawford G, Prince C, Sheikhal Babaei M, Sharp G, *et al.* The EWAS Catalog: a database of epigenome-wide association studies. *Wellcome Open Res* 2022;**7**:41. doi: 10.12688/wellcomeopenres.17598.2
- 47 The EWAS Catalog: a database of epigenome-wide association studies. *Wellcome*
- 48 *Open Res* 2022;**7**:41. doi: 10.12688/wellcomeopenres.17598.2
- 49 42. Li M, Zou D, Li Z, Gao R, Sang J, Zhang Y, *et al.* EWAS Atlas: a curated knowledgebase of epigenome-wide association studies. *Nucleic Acids Res*
- 50 knowledgebase of epigenome-wide association studies. *Nucleic Acids Res*

- 1 2019;**47**:D983-D988. doi: 10.1093/nar/gky1027
- 2 43. Demerath EW, Guan W, Grove ML, Aslibekyan S, Mendelson M, Zhou YH,
3 *et al.* Epigenome-wide association study (EWAS) of BMI, BMI change and waist
4 circumference in African American adults identifies multiple replicated loci. *Hum*
5 *Mol Genet* 2015;**24**:4464-4479. doi: 10.1093/hmg/ddv161
- 6 44. Yamada Y, Horibe H, Oguri M, Sakuma J, Takeuchi I, Yasukochi Y, *et al.*
7 Identification of novel hyper- or hypomethylated CpG sites and genes associated with
8 atherosclerotic plaque using an epigenome-wide association study. *Int J Mol Med*
9 2018;**41**:2724-2732. doi: 10.3892/ijmm.2018.3453
- 10 45. Colicino E, Marioni R, Ward-Caviness C, Gondalia R, Guan W, Chen B, *et al.*
11 Blood DNA methylation sites predict death risk in a longitudinal study of 12, 300
12 individuals. *Aging (Albany NY)* 2020;**12**:14092-14124. doi: 10.18632/aging.103408
- 13 46. Oliva M, Demanelis K, Lu Y, Chernoff M, Jasmine F, Ahsan H, *et al.* DNA
14 methylation QTL mapping across diverse human tissues provides molecular links
15 between genetic variation and complex traits. *Nat Genet* 2023;**55**:112-122. doi:
16 10.1038/s41588-022-01248-z
- 17 47. Wang Y, Song C, Zhao J, Zhang Y, Zhao X, Feng C, *et al.* SEdb 2.0: a
18 comprehensive super-enhancer database of human and mouse. *Nucleic Acids Res*
19 2023;**51**:D280-D290. doi: 10.1093/nar/gkac968
- 20 48. Yang YL, Wu CH, Hsu PF, Chen SC, Huang SS, Chan WL, *et al.* Systemic
21 immune-inflammation index (SII) predicted clinical outcome in patients with
22 coronary artery disease. *Eur J Clin Invest* 2020;**50**:e13230. doi: 10.1111/eci.13230
- 23 49. Hahn J, Bressler J, Domingo-Relloso A, Chen MH, McCartney DL, Teumer A,
24 *et al.* DNA methylation analysis is used to identify novel genetic loci associated with
25 circulating fibrinogen levels in blood. *J Thromb Haemost* 2023. doi:
26 10.1016/j.jtha.2023.01.015
- 27 50. Balta S, Ozturk C. The platelet-lymphocyte ratio: A simple, inexpensive and
28 rapid prognostic marker for cardiovascular events. *Platelets* 2015;**26**:680-681. doi:
29 10.3109/09537104.2014.979340
- 30 51. Lu AT, Quach A, Wilson JG, Reiner AP, Aviv A, Raj K, *et al.* DNA
31 methylation GrimAge strongly predicts lifespan and healthspan. *Aging (Albany NY)*
32 2019;**11**:303-327. doi: 10.18632/aging.101684
- 33 52. Levine ME, Lu AT, Quach A, Chen BH, Assimes TL, Bandinelli S, *et al.* An
34 epigenetic biomarker of aging for lifespan and healthspan. *Aging (Albany NY)*
35 2018;**10**:573-591. doi: 10.18632/aging.101414
- 36 53. Hannum G, Guinney J, Zhao L, Zhang L, Hughes G, Sada S, *et al.* Genome-
37 wide methylation profiles reveal quantitative views of human aging rates. *Mol Cell*
38 2013;**49**:359-367. doi: 10.1016/j.molcel.2012.10.016
- 39 54. Horvath S. DNA methylation age of human tissues and cell types. *Genome*
40 *Biol* 2013;**14**:R115. doi: 10.1186/gb-2013-14-10-r115
- 41 55. Geiler-Samerotte KA, Li S, Lazaris C, Taylor A, Ziv N, Ramjeawan C, *et al.*
42 Extent and context dependence of pleiotropy revealed by high-throughput single-cell
43 phenotyping. *PLoS Biol* 2020;**18**:e3000836. doi: 10.1371/journal.pbio.3000836
- 44 56. van der Stoep M, Korporaal SJ, Van Eck M. High-density lipoprotein as a
45 modulator of platelet and coagulation responses. *Cardiovasc Res* 2014;**103**:362-371.
46 doi: 10.1093/cvr/cvu137
- 47 57. Ridker PM, Bhatt DL, Pradhan AD, Glynn RJ, MacFadyen JG, Nissen SE, *et*
48 *al.* Inflammation and cholesterol as predictors of cardiovascular events among
49 patients receiving statin therapy: a collaborative analysis of three randomised trials.
50 *Lancet* 2023;**401**:1293-1301. doi: 10.1016/S0140-6736(23)00215-5

- 1 58. Shi H, Kichaev G, Pasaniuc B. Contrasting the Genetic Architecture of 30
2 Complex Traits from Summary Association Data. *Am J Hum Genet* 2016;**99**:139-153.
3 doi: 10.1016/j.ajhg.2016.05.013
- 4 59. Fernandez-Sanles A, Sayols-Baixeras S, Subirana I, Senti M, Perez-Fernandez
5 S, de Castro Moura M, *et al.* DNA methylation biomarkers of myocardial infarction
6 and cardiovascular disease. *Clin Epigenetics* 2021;**13**:86. doi: 10.1186/s13148-021-
7 01078-6
- 8 60. Ventham NT, Kennedy NA, Adams AT, Kalla R, Heath S, O'Leary KR, *et al.*
9 Integrative epigenome-wide analysis demonstrates that DNA methylation may
10 mediate genetic risk in inflammatory bowel disease. *Nat Commun* 2016;**7**:13507. doi:
11 10.1038/ncomms13507
- 12 61. Sominen HK, Venkateswaran S, Kilaru V, Marigorta UM, Mo A, Okou DT,
13 *et al.* Blood-Derived DNA Methylation Signatures of Crohn's Disease and Severity of
14 Intestinal Inflammation. *Gastroenterology* 2019;**156**:2254-2265 e2253. doi:
15 10.1053/j.gastro.2019.01.270
- 16 62. Sharp GC, Salas LA, Monnereau C, Allard C, Yousefi P, Everson TM, *et al.*
17 Maternal BMI at the start of pregnancy and offspring epigenome-wide DNA
18 methylation: findings from the pregnancy and childhood epigenetics (PACE)
19 consortium. *Hum Mol Genet* 2017;**26**:4067-4085. doi: 10.1093/hmg/ddx290
- 20 63. Antoun E, Kitaba NT, Titcombe P, Dalrymple KV, Garratt ES, Barton SJ, *et*
21 *al.* Maternal dysglycaemia, changes in the infant's epigenome modified with a diet
22 and physical activity intervention in pregnancy: Secondary analysis of a randomised
23 control trial. *PLoS Med* 2020;**17**:e1003229. doi: 10.1371/journal.pmed.1003229
- 24 64. Wei Y, Cao H, Peng YY, Zhang B. Altered gene expression in dilated
25 cardiomyopathy after left ventricular assist device support by bioinformatics analysis.
26 *Front Cardiovasc Med* 2023;**10**:1013057. doi: 10.3389/fcvm.2023.1013057
- 27 65. Zannas AS, Jia M, Hafner K, Baumert J, Wiechmann T, Pape JC, *et al.*
28 Epigenetic upregulation of FKBP5 by aging and stress contributes to NF-kappaB-
29 driven inflammation and cardiovascular risk. *Proc Natl Acad Sci U S A*
30 2019;**116**:11370-11379. doi: 10.1073/pnas.1816847116
- 31 66. Sakaue S, Kanai M, Tanigawa Y, Karjalainen J, Kurki M, Koshiha S, *et al.* A
32 cross-population atlas of genetic associations for 220 human phenotypes. *Nat Genet*
33 2021;**53**:1415-1424. doi: 10.1038/s41588-021-00931-x
- 34 67. Spracklen CN, Horikoshi M, Kim YJ, Lin K, Bragg F, Moon S, *et al.*
35 Identification of type 2 diabetes loci in 433,540 East Asian individuals. *Nature*
36 2020;**582**:240-245. doi: 10.1038/s41586-020-2263-3
- 37 68. Beunders G, van de Kamp J, Vasudevan P, Morton J, Smets K, Kleefstra T, *et*
38 *al.* A detailed clinical analysis of 13 patients with AUTS2 syndrome further delineates
39 the phenotypic spectrum and underscores the behavioural phenotype. *J Med Genet*
40 2016;**53**:523-532. doi: 10.1136/jmedgenet-2015-103601
- 41 69. Yamauchi T, Hara K, Maeda S, Yasuda K, Takahashi A, Horikoshi M, *et al.* A
42 genome-wide association study in the Japanese population identifies susceptibility
43 loci for type 2 diabetes at UBE2E2 and C2CD4A-C2CD4B. *Nat Genet* 2010;**42**:864-
44 868. doi: 10.1038/ng.660
- 45 70. Noordam R, Sitlani CM, Avery CL, Stewart JD, Gogarten SM, Wiggins KL,
46 *et al.* A genome-wide interaction analysis of tricyclic/tetracyclic antidepressants and
47 RR and QT intervals: a pharmacogenomics study from the Cohorts for Heart and
48 Aging Research in Genomic Epidemiology (CHARGE) consortium. *J Med Genet*
49 2017;**54**:313-323. doi: 10.1136/jmedgenet-2016-104112
- 50 71. Chu AY, Deng X, Fisher VA, Drong A, Zhang Y, Feitosa MF, *et al.*

1 Multiethnic genome-wide meta-analysis of ectopic fat depots identifies loci associated
2 with adipocyte development and differentiation. *Nat Genet* 2017;**49**:125-130. doi:
3 10.1038/ng.3738

4

5

1 **Figure Legends**

2 **Figure 1. Epigenome-wide association studies on DNA methylation and CAD**

3 **adverse outcomes. (a)** Overall study design. Patients in the discovery cohort were

4 enrolled from one medical center, whereas patients in the validation cohort were

5 recruited from three medical centers in the China. Baseline characteristics were

6 collected during the enrollment. DNA methylation of peripheral blood leukocytes was

7 measured by Illumina MethylationEPIC BeadChip on ~ 850,0000 sites. Differentially

8 methylated sites associated with death or MACE were identified, prognostic risk

9 models were built, and biological mechanisms were inferred. This graph was created

10 via <https://www.biorender.com/>. **(b)** EWAS results of death and MACE based on the

11 discovery cohort were presented in Manhattan plot. Red dash line marks the

12 Bonferroni corrected *P* value threshold.

13 **Figure 2. Epigenome-wide association studies of DNA methylation reveals**

14 **differentially methylated CpGs associated with CAD adverse outcomes. (a)**

15 Genomic distribution of the 115 DMPs consistently associated with death and MACE

16 in both the discovery and the validation cohorts. Bars represent *P* values from the

17 discovery cohort. On the bottom of each circle, hypermethylated sites are marked in

18 purple and hypomethylated sites in green. **(b)** Pathway enrichment of the genes

19 connected to DMPs against the KEGG database. **(c)** Diseases or traits enriched among

20 the genes connected to DMPs against the GWASCatalog database. **(d)** Association of

21 the DMPs with three categories of clinical measurements: inflammation indices, lipids,

22 and heart functions. Measurements were recorded in both the discovery (upper) and

23 the validation (lower) cohort. WBC: whole plasma cell count, LMR: lymphocyte-

24 monocyte ratio, NLR: neutrophil-lymphocyte ratio, PLR: platelet-lymphocyte ratio,

25 FIB: fibrinogen, SII: systemic immune-inflammation index, LDLC: low-density

1 lipoprotein cholesterol, HDLC: high-density lipoprotein cholesterol, CHOL: total
2 cholesterol, TRIG: triglycerides, LVEF: left ventricular ejection fraction, LVMI: left
3 ventricular mass index.

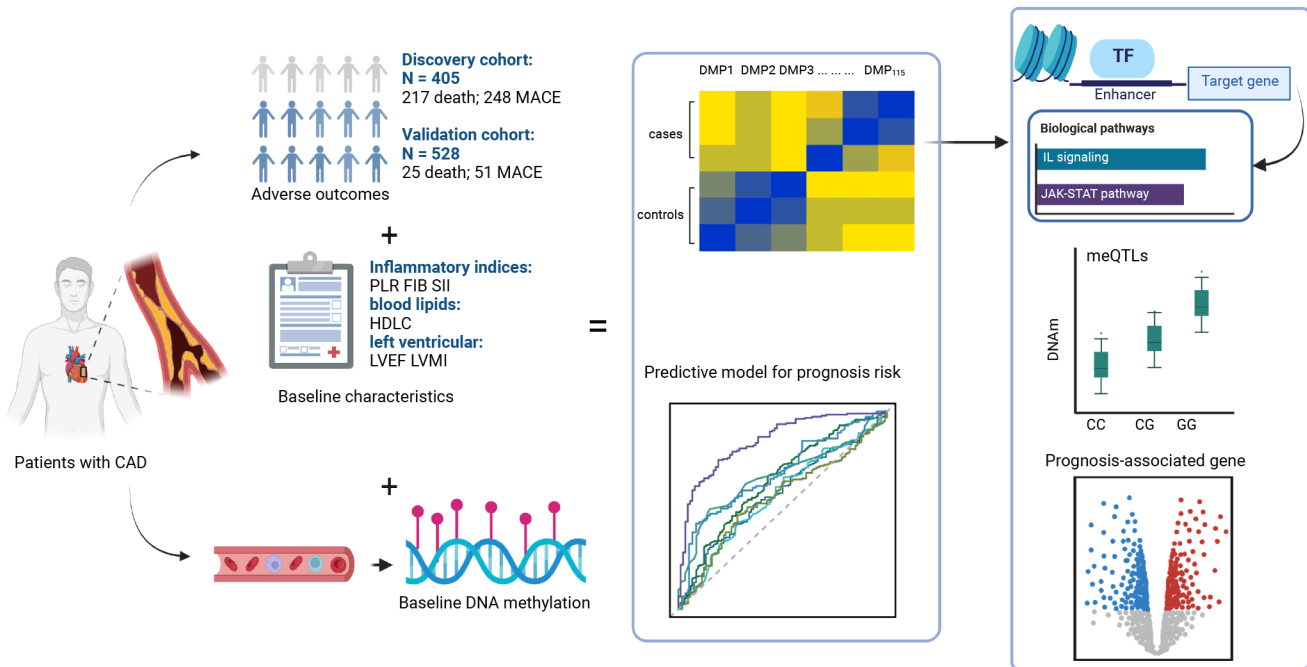
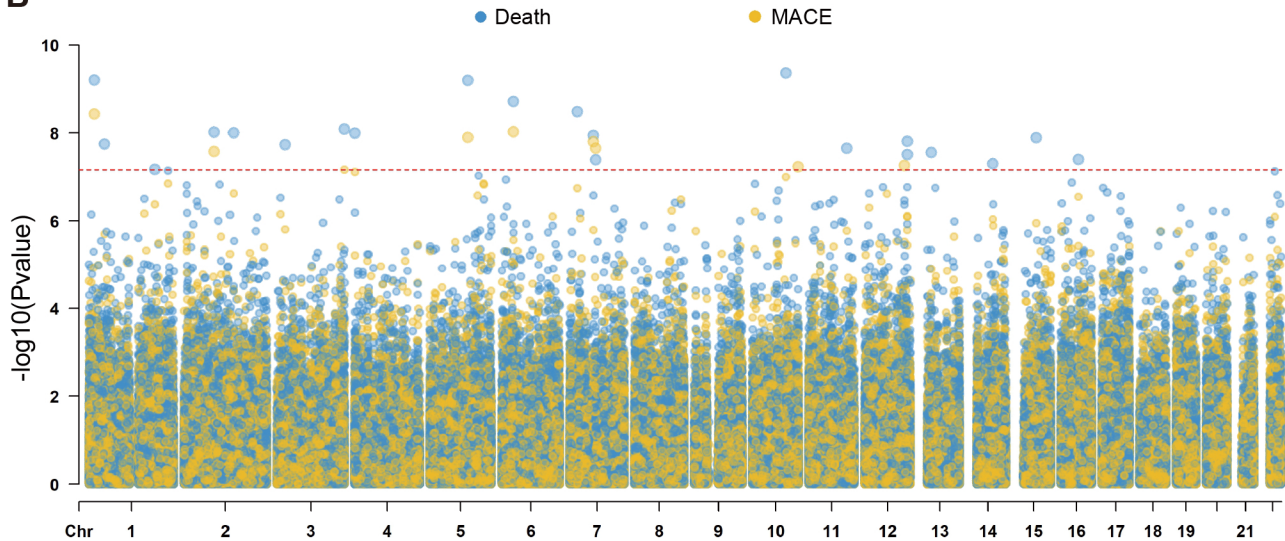
4 **Figure 3. Prognostic models for CAD adverse outcomes. (a-b)** ROC curves of the
5 prognostic models of death constructed from the discovery cohort. The CG model
6 consists of 10 CpG sites. The Ensemble model is composed of CG + Sex + Age + FIB
7 + HDLC + LVEF. **(c)** Predicted risk of death in the validation cohort, by applying the
8 model of CG + Sex + Age. **(d)** Predicted risk of death in the validation cohort, by
9 applying the Ensemble model. **(e-f)** ROC curves of the prognostic models of MACE
10 constructed from the discovery cohort. Features are the same as those for the model of
11 death, except that the CG model consists of 8 CpG sites. **(g)** Predicted risk scores of
12 MACE in the validation cohort, by applying the model of CG + Sex + Age. **(h)**
13 Predicted risk of MACE in the validation cohort, by applying the Ensemble model.
14 PLR: platelet-lymphocyte ratio, FIB: fibrinogen, SII: systemic immune-inflammation
15 index, HDLC: high-density lipoprotein cholesterol, LVEF: left ventricular ejection
16 fraction, LVMI: left ventricular mass index.

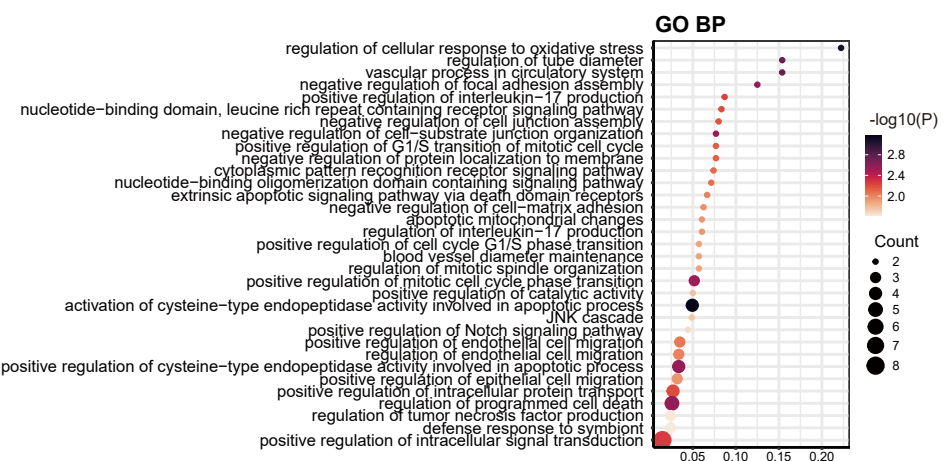
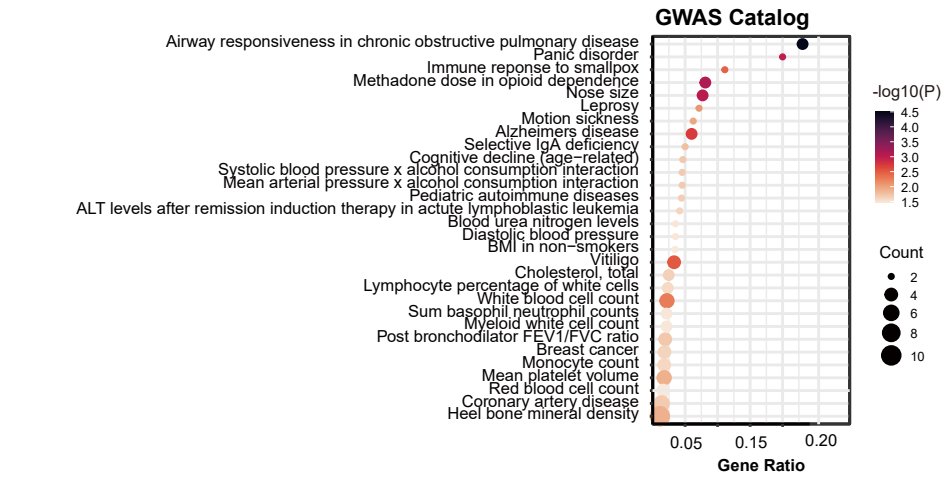
17 **Figure 4. Genetic regulation of DMPs associated with CAD adverse outcomes.**
18 meQTLs associated with the 115 DMPs are listed for **(a)** cis-meQTL and **(b)** trans-
19 meQTL. For each DMP, the number of associated meQTLs was shown. **(c)** The bar
20 plots display the proportion of CpGs having meQTLs. **(d)** Association strength, beta,
21 between the DMPs and the meQTLs. EPIC: all CpGs in the Infinium
22 methylationEPIC Beadchip. DMP: 115 CpGs associated with CAD adverse outcomes.
23 **(e)** Genomic distribution of the 1,842 cis-meQTLs associated with cg20015729,
24 which was located on the *UBE2E2* gene. **(f)** meQTLs for two DMPs located on the
25 promoter of the gene *FKBP5*: cg25563198 associated with 267 cis-meQTLs and 273

1 trans-meQTLs within chromosome 4; cg25114611 associated with 270 trans-meQTLs
2 within chromosome 4. Linkage disequilibrium between the meQTLs, as measured by
3 R^2 , was indicated by the triangle plots.

4 **Figure 5. Expression of the prognostic genes in myocardial infarction (MI) and**
5 **stroke.** Differential gene expression in the MI patients compared to healthy controls
6 in **(a)** MI lesion tissues and **(b)** peripheral blood. **(c)** Expression of the prognostic
7 genes *FKBP5*, *AUTS2* and *UBE2E2* in various cell types in three locations of the MI
8 tissues: myogenic (i.e., nonischemic zone), ischemic, and fibrotic (i.e., advanced MI
9 zone). **(d)** Differential gene expression in the peripheral blood of stroke patients
10 compared to the healthy controls. In all volcano plots of gene expression, prognostic
11 genes inferred from our DNA methylation study were labeled. **(e)** A collection of
12 evidence for the CpG sites associated with the prognostic genes *FKBP5*, *AUTS2*, and
13 *UBE2E2*.

14

A**B**

A**B****C****D**