

**Multi-cohort epigenome-wide association study of cardiovascular disease and all-cause cancer  
incidence: a cardio-oncology approach**

Arce Domingo-Rellosa,<sup>1,2\*</sup> Angela L. Riffo-Campos,<sup>3,4,5\*</sup> Naisi Zhao,<sup>6</sup> Guillermo Ayala,<sup>4</sup> Karin Haack,<sup>7</sup> Carlos Manterola,<sup>3</sup> Dorothy A. Rhoades,<sup>8</sup> Jason G. Umans,<sup>9</sup> M Daniele Fallin,<sup>10</sup> Miguel Herreros-Martinez,<sup>11</sup> Marina Pollan,<sup>2</sup> Eric Boerwinkle,<sup>12</sup> Elizabeth Platz,<sup>13</sup> Miranda Jones,<sup>13</sup> Jan Bressler,<sup>12</sup> Roby Joehanes,<sup>14,15</sup> Daniel Levy,<sup>14,15</sup> Daniel Belsky,<sup>16</sup> Shelley A. Cole,<sup>7</sup> Dominique S. Michaud,<sup>6</sup> Ana Navas-Acien<sup>1</sup> and Maria Tellez-Plaza<sup>2</sup>

<sup>1</sup> Department of Environmental Health Sciences, Columbia University Mailman School of Public Health, New York, NY, USA

<sup>2</sup> Integrative Epidemiology Group, Department of Chronic Diseases Epidemiology, National Center for Epidemiology, Carlos III Health Institute, Madrid, Spain.

<sup>3</sup> Millennium Nucleus on Sociomedicine (SocioMed); and Universidad de La Frontera, Ph.D. Program in Medical Sciences, Temuco, Chile

<sup>4</sup> Department of Computer Science, Universidad de Valencia, Valencia, Spain

<sup>5</sup> Center for Cancer Prevention and Control (CECAN), Santiago, Chile

<sup>6</sup> Department of Public Health & Community Medicine, Tufts University School of Medicine, Boston, MA, USA

<sup>7</sup> Population Health Program, Texas Biomedical Research Institute, San Antonio, TX, USA.

<sup>8</sup> Stephenson Cancer Center, University of Oklahoma Health Sciences Center, Oklahoma City, OK, USA.

<sup>9</sup> MedStar Health Research Institute, Washington DC, USA.

<sup>10</sup> Emory University Rollins School of Public Health, Atlanta, Georgia, USA

<sup>11</sup> Department of Bioinformatics, INCLIVA Biomedical Research Institute, Valencia, Spain

<sup>12</sup> The University of Texas Health Science Center at Houston, Houston, TX, USA

<sup>13</sup> Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA

<sup>14</sup> Population Sciences Branch, National Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda, MD

<sup>15</sup> Framingham Heart Study, Framingham, MA.

<sup>16</sup> Department of Epidemiology, Columbia University Mailman School of Public Health, New York, NY, USA

\*Equal author contribution

**Co-corresponding authors:**

Arce Domingo-Relloso, PhD  
Columbia University  
Mailman School of Public Health  
722 W 168<sup>th</sup> Street  
New York, NY 10032  
E-mail: [ad3531@cumc.columbia.edu](mailto:ad3531@cumc.columbia.edu)

Maria Tellez-Plaza, MD, PhD  
National Center for Epidemiology  
Carlos III Health Institute  
Calle Melchor Fernandez Almagro, 5  
Madrid, Spain, 28029  
E-mail: [m.tellez@isciii.es](mailto:m.tellez@isciii.es)

**Running title:** Common epigenetic signatures for cancer and cardiovascular disease

## **Abstract**

Emerging evidence reveals the complex relationship between cardiovascular disease (CVD) and cancer (Ca), which share common risk factors and biological pathways. We evaluated common epigenetic signatures for CVD and Ca incidence in three ethnically diverse cohorts: the Strong Heart Study (SHS), the Framingham Heart Study and the Atherosclerosis Risk in Communities (separately for European Americans and African Americans). We classified incident cases of only Ca, only CVD and both CVD and Ca (CVD-Ca). Using a Cox proportional hazards model regularized by an elastic-net penalty, we identified 25 common differentially methylated positions (DMPs) across all populations for CVD, six for Ca and seven for CVD-Ca. Of those, nine were statistically significant in a meta-analysis for CVD, one for Ca and two for CVD-Ca. The enrichment analysis of genes annotated to those DMPs pointed to biological pathways involved in Ca and CVD. In the DrugBank database, essential elements and Ca and CVD medications were identified as treatment targets of gene products attributed to overlapping and highly interrelated Ca, CVD and Ca-CVD DMPs. In an additional analysis restricted to the SHS among 950 participants who developed incident CVD, the predictive accuracy for a subsequent incident Ca (N Ca cases=127) increased when adjusting the models for Ca-CVD DMPs. Our results suggest additional molecular pathways and potential treatment targets of precision medicine for CVD and Ca. Also, screening based on common epigenetic signatures of incident CVD and Ca among newly diagnosed CVD patients may help identify individuals at increased Ca risk.

**Key words:** Cardio-oncology, cardiovascular disease, cancer, DNA methylation, multi-cohort.

## **Introduction**

Cardiovascular disease (CVD) and cancer (Ca) are the leading causes of death in the world (1).

While CVD accounts for 40 % of deaths due to chronic diseases, recent reports indicate that Ca has surpassed CVD as leading cause of premature death in high income countries (2). Both diseases share risk factors including aging, cigarette smoking, physical inactivity, unhealthy diet, excess adiposity and genetic and environmental factors (3,4). They also share biological pathways including inflammation, oxidative stress, apoptosis and angiogenesis (3,5,6). Cardotoxicity can follow Ca treatment, such as chemotherapy, radiotherapy (4,5,7,8), targeted Ca therapies and immunotherapy (9). Nevertheless, CVD and Ca have traditionally been studied separately, although the field of cardio-oncology has gained interest in recent years (9).

Epigenetic marks are able to regulate gene expression (10) and their dysregulation might lead to the disruption of essential biological processes, potentially leading to disease (11). However, to our knowledge, epigenetic changes for CVD and Ca have only been studied separately (12,13). For instance, DNA methylation (DNAm), the most well studied epigenetic mark, has not yet been jointly evaluated for CVD and Ca outcomes.

In this study, we identified common epigenetic modifications for CVD and Ca by studying the overlap of differentially methylated positions (DMPs) associated with incident CVD, Ca and both CVD and Ca (CVD-Ca) from separate epigenome-wide association studies (EWAS). We used data from four ethnically diverse cohorts to conduct a multi-cohort study: the Strong Heart Study (SHS) (American Indians), the Framingham Heart Study (FHS) (European Americans) and the Atherosclerosis Risk in Communities study (ARIC) (separately for European Americans and African Americans, respectively denoted as ARICw and ARICb). We subsequently conducted molecular pathway analyses to assess potential downstream biological implications of DMPs associated to both Ca and CVD. These analyses included integrative analysis of protein-protein interaction networks among proteins encoded by genes annotated to identified CVD, Ca and CVD-Ca DMPs, network enrichment analyses, and identification of potential drug targets from the

DrugBank database. In addition, to explore if DNAm might contribute to the identification of candidates for Ca screening in patients with newly diagnosed CVD, we evaluated the predictive ability of DNAm for Ca prediction in participants that had a previous CVD event.

## Methods

### Study populations

*The Strong Heart Study:* The SHS is a prospective cohort study funded by the National Heart, Lung and Blood Institute to investigate CVD and its risk factors in American Indian adults (14). In 1989-1991, a total of 4,549 men and women aged 45–75 years, members of 13 tribes based in the Northern Plains (North and South Dakota), the Southern Plains (Oklahoma), and the Southwest (Arizona), accepted to participate. For this study, one of the tribes declined to participate, leaving 3,517 potential participants. Eligibility criteria for DNAm measurements have been published (15). After exclusion of participants with prevalent CVD, prevalent Ca and missing data in other relevant covariates, 2186 participants were included in this study. Baseline sociodemographic, life-style and anthropometric information were obtained through interviews, physical examinations, and collection of biological samples during the SHS baseline and follow-up visits (14).

*The Framingham Heart Study:* The FHS is a population-based study that started in 1948 (16). 5,209 men and women of European ancestry between the ages of 30 and 62 from the town of Framingham, Massachusetts, were recruited. The FHS Offspring Cohort, founded in 1971, was a second generation study for which children of the Original Cohort were eligible. Spouses were also eligible if they had become pregnant with or sired two or more children by a participant in the Offspring Cohort. DNAm was measured in the Offspring Cohort participants of exam 8 (2005–2008). Clinical variables were recruited in physical examinations and lifestyle interviews. After excluding individuals with prevalent CVD and Ca, as well as individuals with missing data in other

relevant covariates, 1474 participants were included in this study. Sociodemographic and lifestyle factors, as well as anthropometric information, were obtained through interviews, physical examinations and collection of biological samples.

*Atherosclerosis Risk in Communities:* The ARIC study is a prospective cohort study that was initiated in 1987-1989 to investigate cardiovascular disease risk (17). The study involved 15,792 participants sampled from four US communities, including Jackson, MS; Washington County, MD; suburban Minneapolis, MN; and Forsyth County, NC. The baseline cohort consisted of individuals aged 45-64 years, with the Jackson sample comprising exclusively Black residents. Participants underwent a baseline clinical examination (Visit 1) and subsequent follow-up clinical exams (Visits 2–9 completed & Visit 10 planned). DNAm was assessed at Visit 2 (1990-1992) or Visit 3 (1993-1995). For this study, individuals with prevalent cardiovascular disease (CVD) and cancer were excluded, along with those who had missing data for LDL cholesterol, HDL cholesterol, diabetes (yes/no), hypertension treatment (yes/no) and systolic blood pressure. The final analysis included a total of 2212 African American participants and 903 European American participants.

#### Microarray DNA methylation determinations

*The Strong Heart Study:* Biological specimens were collected during a physical exam and were stored at -70°C. DNA from white cells was extracted and stored at the MedStar Health Research Institute laboratory under a strict quality control system. Genomic DNA was bisulfite-converted (Zymo EZ DNA Methylation kits) and DNAm was measured using the Illumina MethylationEPIC BeadChip. Samples were randomized across and within plates to remove potential batch artifacts and confounding effects, and replicate and across-plate control samples were included on every plate.. We conducted sample quality control based on Illumina 850K control probes to assess staining, hybridization, bisulfite conversion, and other parameters. Quality control and

preprocessing were conducted following Illumina's recommendations (15). We determined the total intensity (methylated + unmethylated channel) across all probes measured and any samples with extremely low overall intensity were removed. We ran single sample noob normalization for background correction (18,19), and regression on correlated probes normalization for probe type bias (20). Batch effects for sample row, sample plate and sample isolation time were corrected using the combat package (21). Blood cell counts were calculated using the Flowsorted.Blood.EPIC R package (22), and subsequently used as adjustment variables in the models. Probes in sex chromosomes, cross-hybridizing probes and SNPs with minor allele frequency > 0.05 were excluded (23). We annotated CpGs to the nearest gene (reference genome hg19) using the matchGenes function in *minfi* R package (24).

*The Framingham Heart Study:* In the FHS, buffy coat preparations were obtained from whole-blood samples (Gentra Puregene Blood Kit-Qiagen, Venlo, Netherlands) collected during the eighth examination of the Framingham Offspring Study (2005-2008). DNAm was quantified in the bisulfite converted genomic DNA (EZ DNA Methylation Kit-Zymo Research, Irvine, CA) using Illumina Infinium HumanMethylation450K Beadchip array (25). DASEN methodology in *wateRmelon* R package (26) was used to conduct within laboratory batch normalization of raw data. The exclusion criteria for samples were a missing rate > 1% at detection P-value <0.01, poor matching to the 65 single nucleotide polymorphism (SNP) control probe locations, and identification as outliers using multi-dimensional scaling techniques. In addition, the exclusion criteria for the probes were missing rate > 20% at detection P-value< 0.01, previously identified to map multiple locations, underlying SNP (minor allele frequency > 5% in European ancestry 1000 genomes project data) at the CpG or <10 bp of the single base extension, and location in sex chromosomes.

*Atherosclerosis Risk in Communities:* Genome-wide DNA methylation profiling was conducted using the Illumina Infinium HumanMethylation450K BeadChip array (HM450K) in a total of 2,853 African American participants from the ARIC cohort, covering 483,525 CpG sites (27). In addition, 1,104 European American participants from the same cohort were profiled at 482,815 CpG sites (28). Given that methylation data in the AA and EA participants were measured at two different time points, the current analysis treated the methylation data of African American and European American participants as separate cohorts. Genomic DNA was extracted from peripheral blood leukocyte samples using the Gentra Puregene Blood Kit (Qiagen), and bisulfite conversion of 1 ug genomic DNA was performed using the EZ-96 DNA Methylation Kit (Deep Well Format) (Zymo Research). The bisulfite-converted DNA extracted from peripheral blood leukocytes was then hybridized to the Illumina HumanMethylation450 BeadChip following the Illumina HD Methylation protocol. Participants with a DNA pass rate below 99% were excluded from the analyses, and CpG sites with a detection p-value above 0.01 in more than 5% of the samples were not considered for analysis. Methylation values were processed using the normal exponential out-of-band (NOOB) method for background subtraction and normalized using the Beta MIxture Quantile dilation (BMIQ) method for type I/type II bias correction. Methylation levels were quantified as beta ( $\beta$ ) values, calculated by dividing the intensity of the fluorescence signal from the methylated DNA sequence by the sum of the intensities from both the methylated and unmethylated versions of the sequence. A previous study (27) provided a comprehensive assessment of the DNA methylation measurement, including the evaluation of technical replicates.

#### Cancer and CVD incidence ascertainment and follow-up definition

*The Strong Heart Study:* The SHS uses multiple approaches to assess health history, hospitalizations and vital status over time, including tribal records, death certificates, medical records, and periodic direct contact with participants and their families (14). CVD events were assessed by annual morbidity surveillance reviews of hospitalization and death records through

2017 and at three follow-up SHS research visits conducted in 1993–1995, 1998–1999, and 2014–2016. Incident CVD was defined as the first occurrence of fatal or non-fatal coronary heart disease, stroke or congestive heart failure, or other non-fatal CVD. Details have been described elsewhere (14,29). Ca incidence was assessed by interviews, death certificates and/or chart reviews.

We classified the primary endpoints into three groups: all-cause incident cancer (Ca) in participants that did not develop CVD, all-cause incident CVD in participants that did not develop Ca, and combined incident CVD and cancer CVD (CVD-Ca) for participants who developed both CVD and Ca during the follow-up (in either order).

We calculated follow-up from the date of baseline examination to the date of the Ca or CVD diagnosis, respectively, for Ca and CVD only endpoints, or 31 December 2017 (administrative censoring), whichever occurred first. For the combined CVD-Ca endpoint, we considered the date of the event that happened first. For participants who died of any cause of death other than Ca or CVD, we censored the follow-up at date of death.

*The Framingham Heart Study:* CVD was defined as a composite of coronary heart disease (coronary death, myocardial infarction, coronary insufficiency, and angina), cerebrovascular events (including ischemic stroke, hemorrhagic stroke, and transient ischemic attack), peripheral artery disease (intermittent claudication), and heart failure (30). Medical histories, physical examinations during study visits, hospitalization records and personal physician records were used to identify any possible CVD event. A panel of three experienced investigators reviewed the medical records of suspected new events and made final decision about each event.

Ca incidence was assessed by interviews, death certificates, and/or chart reviews that included pathology reports, and crosschecked with official medical records whenever possible. Ca cases included all cases of primary cancers (malignant or in-situ). Cases described in pathology reports as

benign tumors or tumors of borderline malignancy, cases described as metastases from other primary Ca-s, and basal cell and squamous cell skin Ca cases were excluded (31).

We calculated follow-up from the date of baseline examination to the date of the Ca or CVD diagnosis, respectively, for Ca and CVD only endpoints, or end of follow-up, whichever occurred first. For the combined CVD-Ca endpoint, we considered the date of the event that happened first. For participants who died of any cause of death other than Ca or CVD, we censored the follow-up at date of death.

*Atherosclerosis Risk in Communities:* CVD incidence included incident heart failure (HF), myocardial infarction (MI), ischemic stroke and atrial fibrillation (AF) (32). MI and ischemic stroke were defined based on adjudicated cases using standard ARIC definitions. This involved collecting information through cohort follow-up and active surveillance of hospitalizations. Trained personnel reviewed hospital records, identified relevant International Classification of Diseases, Ninth Revision (ICD-9) codes, and events were reviewed by a committee for adjudication. Incident MI and ischemic stroke events were classified by adjudication committees following ARIC protocols. Incident HF was identified through HF-related hospitalizations during the follow-up period. Incident AF was defined based on ECG findings in subsequent study exams, hospital discharge codes throughout the follow-up, and information from death certificates.

Ca incidence was identified through several methods (33). This included linking the cohort data with state cancer registries in Minnesota, North Carolina, Maryland, and Mississippi. Additionally, active surveillance of the cohort involved recording hospital discharge codes for all participants. Participants self-reported hospitalizations during annual follow-up calls. Any Ca-related hospitalizations not initially identified through registry linkage were considered cases after verification through the retrieval and review of medical records. Information on deaths where cancer was the underlying cause was obtained from death certificates. If a Ca-related death was the

first report of Ca for an individual, it was considered an incident cancer. Follow-up time was calculated from time of blood draw used for the DNA methylation measurements to date of diagnosis (Ca or CVD), last contact date, or December 31, 2019, if last contact date was later than that date.

### Statistical Methods

**Multi-cohort analysis.** To evaluate the consistency of CVD, Ca, and CVD-Ca DMPs across populations, we implemented a multi-step approach. First, we ran GLMnet penalized Cox regression (Cox elastic-net, R package *glmnet*) (34,35) to simultaneously consider all CpGs sites as independent variables in the same model. We fitted an elastic-net Cox model using Ca, CVD and CVD-Ca time-to-event outcomes in separate models. The elastic-net framework is a mix between Ridge and Lasso regression. The algorithm fits a Cox model with penalty controlled by the  $\alpha$  parameter, which can range from 0 – corresponding to Ridge regression to 1 – corresponding to Lasso regression. We selected  $\alpha=0.05$ , an elastic-net choice that works well for methylation data. The regularization parameter  $\lambda$  was selected using 10-folds cross-validation in our study. As all cohorts measured DNAm using the Illumina 450K microarray except the SHS, which used the Illumina MethylationEPIC Beadchip microarray (850K); for this analysis, we restricted the CpGs in the SHS to those present in the 450K array. Second, we found the union set of all DMPs found in all cohorts, and ran targeted elastic-net models restricted to that list of CpGs in all cohorts. Third, we calculated individual hazard ratios (HRs) and 95% confidence intervals using Cox proportional hazards models for those DMPs that were common for all cohorts. Last, we conducted a meta-analysis in those DMPs using the R package *meta* (36). This approach was implemented for each of the three endpoints: CVD, Ca, and CVD-Ca. All models were adjusted for age, sex, BMI, smoking status (never, former, current), DNAm-based smoking score and estimated cell counts. CVD and CVD-Ca models were additionally adjusted for LDL cholesterol, HDL cholesterol, diabetes (yes/no), hypertension treatment (yes/no) and systolic blood pressure. SHS models were

additionally adjusted for study center (Arizona, Oklahoma or North Dakota / South Dakota), five genetic PCs that accounted for population stratification and, for CVD and CVD-Ca models, albuminuria (microalbuminuria, normal albumin levels or macroalbuminuria). FHS models were additionally adjusted for batch, as DNAm data were processed in two different batches. ARIC models additionally adjusted for field center (Jackson, MS; Washington County, MD; suburban Minneapolis, MN; and Forsyth County, NC), and five PCs to correct batch effects in each cohort.

In addition to effect sizes and p-values from the meta-analysis, we reported I<sub>2</sub> statistics, which represent a measure of heterogeneity between studies. I<sub>2</sub> calculates the percentage of variability in the effect sizes which is not caused by sampling error, and is not sensitive to changes in the number of studies included in the meta-analysis (37).

**Molecular pathway analyses.** The DMPs identified after running targeted elastic-net in the union set of DMPs of all cohorts were included in a protein-protein interaction network. Only DMPs associated to protein coding genes were included in the network. We obtained reported biological interactions between the protein nodes from the STRING database v11.5 (38). The STRING database provides a confidence score (from 0 to 1), which estimates the likelihood that an annotated interaction between a pair of proteins is biologically meaningful, specific and reproducible. The automated textmining, experimentally determined and database annotated interactions with a confidence score of 0.4 or greater were included. The network was analyzed and displayed using the yFiles organic layout with Cytoscape v3.9.1 (39). A network enrichment analysis was performed by incorporating available information for the relationships between nodes based on Gene Ontology (GO) Biological Process, GO Molecular Function, and UniProt databases through a combination of methods that include multiple testing correction, two-sided Kolmogorov-Smirnov test, and hierarchical clustering of the STRING network itself (38). We also attempted to identify common relevant biological mechanisms for Ca and CVD by evaluating which nodes are drug targets within DrugBank database v5.1.10 DrugBank combines detailed data annotations with

comprehensive drug target information (40), which enables assessing if gene-products are potential drug targets of specific targeting compounds. Among the 5,225 non-redundant proteins corresponding to drug entries and associated drug targets in DrugBank, we searched for those related to the common nodes from overlapping Ca, CVD and CVD-Ca DMPs.

*Predictive models of incident cancer in participants who developed cardiovascular disease first in the Strong Heart Study.* To explore if DNAm might contribute to the identification of candidates for Ca screening in patients with newly diagnosed CVD, we evaluated the predictive ability of baseline blood DNAm signatures, as measured by the concordance index (C index), for Ca prediction, comparing different sets of candidate DMPs (i.e. identified DMPs and relevant protein network nodes) in progressively adjusted models. We restricted this analysis to participants who had a CVD event at any time and excluded individuals that developed Ca before CVD, leaving N=950 individuals for these analyses. The reverse analysis, evaluating the predictive ability of blood DNAm for CVD among individuals who had Ca first was not feasible due to the small number of cases (N= 15 Ca cases happening before CVD). We conducted four progressively adjusted models: 1) adjusted only for Ca and CVD risk factors (no CpGs) (Model 1); 2) Model 1 further adjusted for the identified Ca DMPs (Supplementary File 1, sheet A); 3) Model 1 further adjusted for the identified CVD-Ca DMPs (Supplementary File 1, sheet C); 4) Model 1 further adjusted for the 265 CpGs annotated to the 69 protein nodes from any combination of overlapping Ca, CVD and CVD-Ca DMPs (green nodes in Figure 2, Supplementary File 1, sheet N).

## Results

*Descriptive analysis.* In the SHS, we observed 277, 823, and 142 incident Ca, CVD, and CVD-Ca cases, respectively (Table 1). In the FHS, we observed 222, 337 and 51 incident CVD, cancer and CVD-Ca cases, respectively. In ARICb, we observed 389 incident cases of Ca, 622 incident cases of CVD, and 244 incident cases of CVD-Ca. Additionally, in ARICw, there were 193 incident cases

of cancer, 222 incident cases of CVD, and 84 incident cases of CVD-Ca. In all cohorts, participants with incident CVD were older and more likely to have baseline hypertension and diabetes compared to non-cases. Baseline current smoking was more common in participants with incident Ca, CVD and CVD-Ca (Table 1), compared to non-cases.

**Multi-cohort analysis.** A summary of the number of DMPs identified for each cohort and each endpoint, as well as the number of significant DMPs identified in the meta-analysis, can be found in Figure 1. The overlaps between DMPs across cohorts for each endpoint can be found in Figure S1. 25 DMPs were common across all populations for CVD, six for Ca and seven for CVD-Ca. Nine DMPs were statistically significant in the meta-analysis for CVD (annotated to genes *WDR37*, *TXNIP*, *ISOC2*, *SERINC5*, *ADAM33*, *EHMT1*, *LINC01257*, *LOX*, *TBXT*) at a p-value cut-off of 0.1. One DMP was significant for Ca (annotated to *TENM2*), and two were significant for CVD-Ca (annotated to *CNPY1* and *CDH15*) (Table 2).

**Molecular pathways analyses.** In protein-protein interaction network analysis, only DMP-associated genes reported in common for at least three cohorts were included. Of these 477 unique genes, 165 ncRNA or protein-coding genes without connections in the network were discarded. As a result, the protein interaction network contained 312 nodes and 574 edges (Figure 2, Supplementary File 1 sheets M and N). The most connected node in the network was HDAC4 with 23 edges, associated to Ca in ARICw, FHS and SHS, and CVD in ARICb, ARICw and SHS. The next most connected nodes were SMAD3, H4-16 and GLI2 with 20, 20 and 19 interactions, respectively, and associated only to CVD DMPs. The most connected nodes associated only to CVD-Ca DMPs in the network (nodes displaying > 10 connections) were PAX6 and DNMT3A with 15 and 11 interactions respectively. The most connected nodes associated only to Ca DMPs in the network were YY1 and NCOR2 with 15 and 11 interactions respectively. The 69 overlapping nodes between at least two of Ca, CVD and CVD-Ca related DMPs included NKX2-5 (16 edges) and GNG7 (11 edges)

associated to CVD and Ca, and Ca and CVD-Ca DMPs, respectively (Figure 2). MGMT, PTPRN2 and PRDM16 were associated to CVD-Ca, Ca and CVD DMPs in at least 2 cohorts, with 8 interactions.

**Network enrichment analysis.** A total of 92 GO and 12 UniProt terms were reported ( $FDR < 0.05$ ). The top five *GO Biological Process* were multicellular organism development (GO:0007275,  $FDR = 1.83e-08$ ), system development (GO:0048731,  $FDR = 1.83e-08$ ), anatomical structure development (GO:0048856,  $FDR = 1.83e-08$ ), nervous system development (GO:0007399,  $FDR = 7.99e-08$ ) and developmental process (GO:0032502,  $FDR = 7.99e-08$ ); in *Molecular Function* ( $FDR = 0.0035$ ) were transcription factor binding (GO:0008134), activating transcription factor binding (GO:0033613), sequence-specific DNA binding (GO:0043565), molecular function regulator (GO:0098772) and transcription regulator activity (GO:0140110); and in *UniProt* were disease (KW-9995,  $FDR = 5.80e-06$ ), repressor (KW-0678,  $FDR = 0.00014$ ), phosphoprotein (KW-0597,  $FDR = 0.00031$ ), disease mutation (KW-0225,  $FDR = 0.00086$ ) and developmental protein (KW-0217,  $FDR = 0.0013$ ) (Supplementary File 1, O sheet). Additionally, in the DrugBank database we identified drug target proteins (UniProt ID) related to nutritional unbalance and diarrhea (MGMT [P16455], HDAC4 [P56524], and RBP7 [Q96R05]); cancer treatment (PTK6 [Q13882] and HDAC4 [P56524]); and cardiometabolic diseases (CPT1A[P50416] and PTK6 [Q13882]) (Table 4).

**Incident cancer predictive accuracy among individuals with a prior cardiovascular event in the Strong Heart Study.** 127 individuals developed Ca after an initial CVD event (accumulated follow-up time was 9185.2 person-years). The C index for the fully adjusted model without blood DNAm was 0.65 for all-cancer (Table S1). The best predictive ability for a cancer event following a CVD event was observed when including the 380 Ca-CVD DMPs (C index = 0.96).

## Discussion

In this multi-cohort study conducted across four ethnically diverse populations, we found 25 common DMPs across all populations for CVD, six for Ca and seven for Ca-CVD. Of those, nine were statistically significant in a meta-analysis for CVD, one for Ca and two for Ca-CVD. The proteins encoded by those genes were involved in common molecular pathways for Ca and CVD. Consistently, in a protein network enrichment analysis, nodes related to widely known mechanisms for both diseases were interconnected (transcription factor binding [GO:0008134], disease [KW-9995] or multicellular organism development [GO:0007275], which include nodes like HDAC4, SMAD3 or TBX3), further supporting common underlying pathways. The fact that many DMPs were not common across populations might indicate part of the epigenomic signature for Ca-CVD to be population-specific. In restricted analyses among participants who initially had an incident CVD event, we observed a markedly strong increase in the predictive accuracy for incident Ca after including baseline methylation data of Ca, CVD and Ca-CVD DMPs.

Most of the genes annotated to the DMPs identified in our meta-analysis have biological functions relevant for CVD and Ca. Among the CVD DMPs, the *WR37* gene plays a role in cell cycle progression, signal transduction, apoptosis, and gene regulation. Missense variants in *WDR37* in humans have been causally associated with a multisystemic disorder characterized by anomalies of the cardiovascular system, among others (41). In addition, the *TXNIP* gene is an oxidative and inflammation-related marker which has been identified as a potential biomarker for CVD, as well as a target for preventive and curative medicine (42,43). Also, increasing evidence has shown that *TXNIP* might act as a tumor suppressor gene in several Ca types such as liver, breast and lung Ca (44). The *EHMT1* gene is a dimethyltransferase that has been identified as a potential therapeutic approach for the treatment of heart disease (45). It also plays a critical role in the regulation of Ca cell apoptosis and the cell cycle (46). The *LOX* gene plays a role on the synthesis of extracellular matrix (47). Since the cardiovascular system performance is influenced by the structure and composition of the extracellular matrix, dysregulations in *LOX* expression might as well contribute

to CVD progression. Some studies have also reported a tumor-suppressor role of *LOX* (48). The fact that several genes identified in our CVD meta-analysis have biological functions associated with Ca in addition to CVD highlights the fact that these two diseases share many common biological pathways.

The only Ca DMP identified in our meta-analysis was the *TENM1* gene, from the teneurins gene family. It might be involved in Ca initiation, progression and drug resistance. Dysregulations in teneurins expression have been associated with several tumor types and patient survival, and have also been proposed as molecular diagnostic and prognostic biomarkers and as targets for Ca treatments (49). One of the CVD-Ca DMPs identified in our meta-analysis is annotated to the *CDH15* gene, from the cadherin family. Methylated *CDH15* has been associated with an increased risk of liver Ca (50). In addition, several cadherins are regulated during the epithelial-to-mesenchimal transition, which is essential for metastases (51).

In the protein-protein interaction network, the most connected nodes commonly associated with Ca, CVD and Ca-CVD DMPs included the *PTPRN2*, *MGMT* and *PRDM16* genes (Figure 2). *PTPRN2* has known functions in both Ca and CVD and might encode a phosphatidylinositol phosphatase. DNA methylation in this gene was associated with vascular and cardiac disease (52) and was also included in a diagnostic panel for squamous cell lung Ca (53). DNA methylation of *MGMT* and *PRDM16* genes has a known role in Ca, however, to our knowledge, this is the first study reporting an association with CVD. The *MGMT* gene encodes a DNA repair protein, which repairs the toxic lesion produced by alkylating agents. DNA methylation of the *MGMT* gene has been associated with many types of Ca, for instance, as a prognostic biomarker in glioblastoma (54), as biomarker for detecting field cancerization in breast cancer (55), and associated with gastric cancer risk (56). The *PRDM16* gene encodes a zinc finger transcription factor and contains an N-termR domain. DNA methylation of this gene has been involved in non-small cell lung cancer (57) and pediatric acute myeloid leukemia (58), among others. The hub of the protein-protein interaction network (the most connected node) was *HDAC4* (Figure 2). The proteic product of this gene possesses histone deacetylase activity

widely associated to Ca (59), but also CVD (60). The *H4-16* and *SMAD3* genes were the second most connected nodes in the network, both associated only with CVD in our study, and also associated with Ca in other studies (61,62).

The network enrichment analyses identified general pathways such as developmental process, transcription factor binding or disease, associated with a large number of chronic conditions such as liver diseases, diabetes mellitus or eating disorders, in addition to Ca and CVD. The DrugBank search showed proteins encoded by genes annotated to Ca-CVD DMPs as targets for Ca and cardiometabolic risk medications, as well as essential nutrients such as zinc and cysteine metabolites. Zinc supplementation has now been questioned given recent evidence suggesting high zinc exposure might be associated with increased oxidative stress and CVD risk, as well as potential deficiencies in the study designs that support zinc supplementation (63–65). Interestingly, non-essential divalent metals that compete for zinc and cysteine binding sites, such as cadmium (66), have been related to cancer and CVD in the SHS and other populations (67–71).

In this study, we evaluated the epigenetic susceptibility of CVD patients to develop Ca. Experimental data support that CVD can predispose to Ca development. In a mouse model, heart failure stimulated intestinal tumor load, with the severity of left ventricular dysfunction and fibrotic scar being strongly correlated with tumor growth (70). In humans, elevated cardiac and inflammation biomarkers in apparently healthy individuals were predictive of new-onset Ca independently of Ca risk factors (70). Alternatively, Ca patients are susceptible to suffer from stroke and HF complications (71,72). Lung Ca survivors have a higher risk of CHD and ischemic stroke compared with the non-Ca population (73). We were not able to evaluate the susceptibility of Ca patients to develop CVD due to the small number of Ca cases that subsequently developed CVD. Further studies focused on DNAm profiles of individuals that develop CVD after Ca are needed.

Of note, we found many more common significant DMPs for CVD than for Ca. This could be related to the heterogeneity of the Ca endpoint, as epigenetic changes might be specific to different types of Ca, which constitutes one of the limitations of our study. We were not able to

conduct separate analysis for each type of specific cancer due to the small number of cases.

Additional onco-cardiology studies are needed focusing on specific CVD and Ca endpoints. In addition, non-fatal cancer data in the SHS might be incomplete, as the SHS is not linked to Ca registry data, and incident Ca identification is dependent on self-report and SHS review of medical records through a surveillance system that was initially developed for CVD outcomes. However, the multi-cohort approach and the ethnical diversity of the participants included in this study adds robustness to the identified DMPs, and FHS and ARIC do have linkage to the Ca registry data. In addition, incident Ca, CVD and Ca-CVD endpoints were analyzed using an informative elastic-net method, which enabled avoiding severe multiple-comparison correction methods, while accounting for potential correlations between CpGs. Other strengths of our study include the integration of data from multiple diverse cohorts, the ability to account for a wide range of confounders, the prospective nature and long follow-up of the studies and the combination of an EWAS with subsequent bioinformatics analyses to evaluate the biological plausibility of the findings.

## **Conclusions**

We identified novel, and also replicated previously reported epigenetic signatures for Ca and CVD. Our bioinformatic analyses additionally support that underlying common molecular pathways are related to Ca and CVD onset. Future studies that experimentally evaluate the role of the identified gene targets in CVD and Ca are needed. The common epigenomic signatures for Ca and CVD found in this study could potentially contribute to identify newly diagnosed CVD individuals at increased Ca risk, thus enabling the precise prevention and control of Ca and CVD.

## **Sources of Funding**

This work was supported by grants of the National Heart, Lung, and Blood Institute (NHLBI) (under contract numbers 75N92019D00027, 75N92019D00028, 75N92019D00029, &

75N92019D00030) and previous grants (R01HL090863, R01HL109315, R01HL109301, R01HL109284, R01HL109282, and R01HL109319 and cooperative agreements U01HL41642, U01HL41652, U01HL41654, U01HL65520, and U01HL65521), by the National Institute of Health Sciences (R01ES021367, R01ES025216, P42ES033719, P30ES009089), by the Spanish Funds for Research In Health Sciences, Carlos III Health Institute, co-funded by European Regional Development Fund (CP12/03080 and PI15/00071), and Spanish Agency for Research (PID2019-108973RB-C21 and PID2020-117114GB-I00) by “Ministerio de Ciencia e Innovación”, María Zambrano Nº ZA21-063 grant founded by the Ministry of Universities of the Government of Spain, financed by the European Union, NextGeneration EU, to A.R.-C.; ANID–Millennium Science Initiative Program—NCS2021\_013 and ANID FONDAP 152220002 (CECAN) and a fellowship from “La Caixa” Foundation (ID 100010434), code “LCF/BQ/DR19/11740016”.

The Atherosclerosis Risk in Communities study has been funded in whole or in part with Federal funds from the National Heart, Lung, and Blood Institute, National Institutes of Health, Department of Health and Human Services, under Contract nos. (75N92022D00001, 75N92022D00002, 75N92022D00003, 75N92022D00004, 75N92022D00005). The authors thank the staff and participants of the ARIC study for their important contributions.

ARIC Cancer Studies are also supported by the National Cancer Institute (U01 CA164975) and the National Heart, Lung, and Blood Institute, National Institutes of Health, Department of Health and Human Services, under Contract nos. (75N92022D00001, 75N92022D00002, 75N92022D00003, 75N92022D00004, 75N92022D00005). Cancer data was provided by the Maryland Cancer Registry, Center for Cancer Prevention and Control, Maryland Department of Health, with funding from the State of Maryland and the Maryland Cigarette Restitution Fund. The collection and availability of cancer registry data is also supported by the Cooperative Agreement NU58DP006333, funded by the Centers for Disease Control and Prevention.

The Framingham Heart Study (FHS) is funded by National Institutes of Health contract N01-HC-25195. The laboratory work for this investigation was funded by the Division of Intramural

Research, National Heart, Lung, and Blood Institutes, National Institutes of Health and an NIH Director's Challenge Award (D. Levy, Principal Investigator).

The content of this work is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health, Carlos III Health Institute, Centers for Disease Control and Prevention or the Department of Health and Human Services.

## **Disclosures**

None

## **References**

1. Michel L, Schadendorf D, Rassaf T. Oncocardiology: new challenges, new opportunities [Internet]. Vol. 45, Herz. Springer Medizin; 2020 [cited 2021 May 1]. p. 619–25. Available from: <https://doi.org/10.1007/s00059-020-04951-x>
2. Mahase E. Cancer overtakes CVD to become leading cause of death in high income countries. BMJ [Internet]. 2019 Sep 3 [cited 2023 Apr 28];366. Available from: <https://www.bmjjournals.org/content/366/bmj.l5368>
3. Koene RJ, Prizment AE, Blaes A, Konety SH. Shared risk factors in cardiovascular disease and cancer. Circulation [Internet]. 2016 [cited 2021 May 1];133(11):1104–14. Available from: <https://pubmed.ncbi.nlm.nih.gov/26976915/>
4. Al-Kindi SG, Oliveira GH. Onco-Cardiology: A Tale of Interplay Between 2 Families of Diseases [Internet]. Vol. 91, Mayo Clinic Proceedings. Elsevier Ltd; 2016 [cited 2021 May 1]. p. 1675–7. Available from: <https://pubmed.ncbi.nlm.nih.gov/27916153/>
5. Giza DE, Iliescu G, Hassan S, Marmagkiolis K, Iliescu C. Cancer as a Risk Factor for Cardiovascular Disease [Internet]. Vol. 19, Current Oncology Reports. Current Medicine Group LLC 1; 2017 [cited 2021 May 1]. Available from: <https://pubmed.ncbi.nlm.nih.gov/28421481/>
6. Tapia-Vieyra JV, Delgado-Coello B, Mas-Oliva J. Atherosclerosis and Cancer; A Resemblance with Far-reaching Implications. Arch Med Res [Internet]. 2017 Jan 1 [cited 2023 Apr 22];48(1):12–26. Available from: <https://pubmed.ncbi.nlm.nih.gov/28577865/>
7. Hong RA, Iimura T, Sumida KN, Eager RM. Cardio-Oncology/Onco-Cardiology. Clin Cardiol [Internet]. 2010 Dec 1 [cited 2021 May 1];33(12):733–7. Available from: <http://doi.wiley.com/10.1002/clc.20823>
8. Aleman BMP, Moser EC, Nuver J, Suter TM, Maraldo M V., Specht L, et al. Cardiovascular disease after cancer therapy. Eur J Cancer, Suppl. 2014;12(1):18–28.
9. Michel L, Schadendorf D, Rassaf T. Oncocardiology: new challenges, new opportunities. Herz [Internet]. 2020 Nov 1 [cited 2023 Apr 22];45(7):619–25. Available from:

- https://link.springer.com/article/10.1007/s00059-020-04951-x
10. Allis CD, Jenuwein T. The molecular hallmarks of epigenetic control. *Nat Rev Genet.* 2016 Aug 1;17(8):487–500.
  11. Zoghbi HY, Beaudet AL. Epigenetics and Human Disease. *Cold Spring Harb Perspect Biol [Internet].* 2016 Feb 1 [cited 2023 Apr 28];8(2):1–28. Available from: <https://pubmed.ncbi.nlm.nih.gov/26834142/>
  12. Kinnaird A, Zhao S, Wellen KE, Michelakis ED. Metabolic control of epigenetics in cancer. *Nat Rev Cancer.* 2016 Oct 24;16(11):694–707.
  13. van der Harst P, de Windt LJ, Chambers JC. Translational Perspective on Epigenetics in Cardiovascular Disease [Internet]. Vol. 70, *Journal of the American College of Cardiology.* Elsevier USA; 2017 [cited 2021 May 1]. p. 590–606. Available from: <https://pubmed.ncbi.nlm.nih.gov/28750703/>
  14. Lee ET, Welty TK, Fabsitz R, Cowan LD, Le NA, Oopik AJ, et al. The Strong Heart Study. A study of cardiovascular disease in American Indians: design and methods. *Am J Epidemiol [Internet].* 1990 Dec [cited 2018 Apr 5];132(6):1141–55. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/2260546>
  15. Domingo-Reloso A, Riffo-Campos AL, Haack K, Rentero-Garrido P, Ladd-Acosta C, Fallin DM, et al. Cadmium, Smoking, and Human Blood DNA Methylation Profiles in Adults from the Strong Heart Study. *Environ Health Perspect [Internet].* 2020 [cited 2023 Apr 22];128(6). Available from: <https://pubmed.ncbi.nlm.nih.gov/32484362/>
  16. Tsao CW, Vasan RS. Cohort Profile: The Framingham Heart Study (FHS): overview of milestones in cardiovascular epidemiology. *Int J Epidemiol [Internet].* 2015 Dec 1 [cited 2022 Feb 14];44(6):1800–13. Available from: <https://pubmed.ncbi.nlm.nih.gov/26705418/>
  17. Wright JD, Folsom AR, Coresh J, Sharrett AR, Couper D, Wagenknecht LE, et al. The Atherosclerosis Risk in Communities (ARIC) study: JACC Focus Seminar 3/8. *J Am Coll Cardiol [Internet].* 2021 Jun 6 [cited 2023 Jun 1];77(23):2939. Available from: [/pmc/articles/PMC8667593/](https://pmc/articles/PMC8667593/)
  18. Fortin J-P, Triche TJ, Hansen KD. Preprocessing, normalization and integration of the Illumina HumanMethylationEPIC array with minfi. *Bioinformatics [Internet].* 2016 Dec 28 [cited 2018 Mar 30];33(4):btw691. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/28035024>
  19. Triche TJ, Weisenberger DJ, Van Den Berg D, Laird PW, Siegmund KD. Low-level processing of Illumina Infinium DNA Methylation BeadArrays. *Nucleic Acids Res [Internet].* 2013 Apr [cited 2018 Mar 30];41(7):e90–e90. Available from: <https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gkt090>
  20. Niu L, Xu Z, Taylor JA. RCP: a novel probe design bias correction method for Illumina Methylation BeadChip. *Bioinformatics.* 2016;32(17):2659–63.
  21. Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD. The SVA package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics [Internet].* 2012 Mar [cited 2021 May 23];28(6):882–3. Available from: <https://pubmed.ncbi.nlm.nih.gov/22257669/>
  22. Salas LA, Koestler DC, Butler RA, Hansen HM, Wiencke JK, Kelsey KT, et al. An optimized library for reference-based deconvolution of whole-blood biospecimens assayed using the Illumina HumanMethylationEPIC BeadArray. *Genome Biol [Internet].* 2018 May 29 [cited 2023 Apr 28];19(1). Available from: <https://pubmed.ncbi.nlm.nih.gov/29843789/>

23. McCartney DL, Walker RM, Morris SW, McIntosh AM, Porteous DJ, Evans KL. Identification of polymorphic and off-target probe binding sites on the Illumina Infinium MethylationEPIC BeadChip. *Genomics* data [Internet]. 2016 Sep [cited 2018 Jun 8];9:22–4. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/27330998>
24. Doi A, Park I-H, Wen B, Murakami P, Aryee MJ, Irizarry R, et al. Differential methylation of tissue- and cancer-specific CpG island shores distinguishes human induced pluripotent stem cells, embryonic stem cells and fibroblasts. *Nat Genet* [Internet]. 2009 Dec [cited 2019 Mar 12];41(12):1350–3. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/19881528>
25. Joehanes R, Just AC, Marioni RE, Pilling LC, Reynolds LM, Mandaviya PR, et al. Epigenetic Signatures of Cigarette Smoking. *Circ Cardiovasc Genet* [Internet]. 2016 Oct [cited 2019 Mar 14];9(5):436–47. Available from: <https://www.ahajournals.org/doi/10.1161/CIRGENETICS.116.001506>
26. Pidsley R, Y Wong CC, Volta M, Lunnon K, Mill J, Schalkwyk LC. A data-driven approach to preprocessing Illumina 450K methylation array data. *BMC Genomics* [Internet]. 2013 May 1 [cited 2023 Apr 28];14(1):1–10. Available from: <https://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2164-14-293>
27. Bose M, Wu C, Pankow JS, Demerath EW, Bressler J, Fornage M, et al. Evaluation of microarray-based DNA methylation measurement using technical replicates: The atherosclerosis risk in communities (ARIC) study. *BMC Bioinformatics* [Internet]. 2014 Sep 19 [cited 2023 Jun 1];15(1):1–10. Available from: <https://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-15-312>
28. Huan T, Joehanes R, Song C, Peng F, Guo Y, Mendelson M, et al. Genome-wide identification of DNA methylation QTLs in whole blood highlights pathways for cardiovascular disease. *Nat Commun* 2019 101 [Internet]. 2019 Sep 19 [cited 2023 Jun 15];10(1):1–14. Available from: <https://www.nature.com/articles/s41467-019-12228-z>
29. Strongheart Study - Center for American Indian Health Research - College of Public Health [Internet]. [cited 2023 Apr 28]. Available from: <https://strongheartstudy.org/>
30. D'Agostino RB, Vasan RS, Pencina MJ, Wolf PA, Cobain M, Massaro JM, et al. General cardiovascular risk profile for use in primary care: the Framingham Heart Study. *Circulation* [Internet]. 2008 Feb [cited 2021 Dec 15];117(6):743–53. Available from: <https://pubmed.ncbi.nlm.nih.gov/18212285/>
31. BE K, GL S, A S. The Cancer Experience in the Framingham Heart Study Cohort. *Cancer* [Internet]. 1991 [cited 2020 Jun 20];67(1). Available from: <https://pubmed.ncbi.nlm.nih.gov/1845934/>
32. Almuwaqqat Z, Jokhadar M, Norby FL, Lutsey PL, O'Neal WT, Seyerle A, et al. Association of Antidepressant Medication Type With the Incidence of Cardiovascular Disease in the ARIC Study. *J Am Hear Assoc Cardiovasc Cerebrovasc Dis* [Internet]. 2019 Jun 6 [cited 2023 Jun 1];8(11). Available from: [/pmc/articles/PMC6585369/](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6585369/)
33. Joshu CE, Barber JR, Coresh J, Couper DJ, Mosley TH, Vitolins MZ, et al. Enhancing the Infrastructure of the Atherosclerosis Risk in Communities (ARIC) Study for Cancer Epidemiology Research: ARIC Cancer. *Cancer Epidemiol Biomarkers Prev* [Internet]. 2018 Mar 1 [cited 2023 Jun 1];27(3):295. Available from: [/pmc/articles/PMC5835193/](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5835193/)
34. Friedman J, Hastie T, Tibshirani R, Narasimhan B, Simon N, Qian J, et al. Package “glmnet” Type Package Title Lasso and Elastic-Net Regularized Generalized Linear Models NeedsCompilation yes. 2019;
35. Simon N, Friedman J, Hastie T, Tibshirani R. Regularization Paths for Cox's Proportional

- Hazards Model via Coordinate Descent. *J Stat Softw* [Internet]. 2011 Mar 9 [cited 2023 Apr 28];39(5):1–13. Available from: <https://www.jstatsoft.org/index.php/jss/article/view/v039i05>
- 36. Schwarzer G, Carpenter JR, Rücker G. Meta-Analysis with R. 2015 [cited 2023 Apr 28]; Available from: <https://link.springer.com/10.1007/978-3-319-21416-0>
  - 37. 7.1 Heterogeneity statistics | Doing Meta-Analysis in R and exploring heterogeneity using metaforest [Internet]. [cited 2023 Apr 30]. Available from: <https://cjvanlissa.github.io/Doing-Meta-Analysis-in-R/heterogeneity-statistics.html>
  - 38. Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, et al. STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* 2019 Jan;47(D1):D607–13.
  - 39. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: A software Environment for integrated models of biomolecular interaction networks. *Genome Res.* 2003 Nov;13(11):2498–504.
  - 40. Knox C, Law V, Jewison T, Liu P, Ly S, Frolkis A, et al. DrugBank 3.0: a comprehensive resource for “omics” research on drugs. *Nucleic Acids Res* [Internet]. 2011 Jan [cited 2023 Apr 28];39(Database issue). Available from: <https://pubmed.ncbi.nlm.nih.gov/21059682/>
  - 41. Reis LM, Sorokina EA, Thompson S, Muheisen S, Velinov M, Zamora C, et al. De Novo Missense Variants in WDR37 Cause a Severe Multisystemic Syndrome. *Am J Hum Genet* [Internet]. 2019 Aug 8 [cited 2023 Apr 22];105(2):425. Available from: [/pmc/articles/PMC6698968/](https://pmc/articles/PMC6698968/)
  - 42. Domingues A, Jolibois J, de Rougé PM, Nivet-Antoine V. The Emerging Role of TXNIP in Ischemic and Cardiovascular Diseases; A Novel Marker and Therapeutic Target. *Int J Mol Sci* 2021, Vol 22, Page 1693 [Internet]. 2021 Feb 8 [cited 2023 Apr 28];22(4):1693. Available from: <https://www.mdpi.com/1422-0067/22/4/1693/htm>
  - 43. Domingo-Rellosa A, Makhani K, Riffo-Campos AL, Tellez-Plaza M, Klein KO, Subedi P, et al. Arsenic Exposure, Blood DNA Methylation, and Cardiovascular Disease. *Circ Res* [Internet]. 2022 Jul 8 [cited 2023 Jun 19];131(2):E51–69. Available from: <https://www.ahajournals.org/doi/abs/10.1161/CIRCRESAHA.122.320991>
  - 44. Chen Y, Ning J, Cao W, Wang S, Du T, Jiang J, et al. Research Progress of TXNIP as a Tumor Suppressor Gene Participating in the Metabolic Reprogramming and Oxidative Stress of Cancer Cells in Various Cancers. *Front Oncol* [Internet]. 2020 Oct 21 [cited 2023 Apr 22];10:568574. Available from: [/pmc/articles/PMC7609813/](https://pmc/articles/PMC7609813/)
  - 45. Thienpont B, Aronsen JM, Robinson EL, Okkenhaug H, Loche E, Ferrini A, et al. The H3K9 dimethyltransferases EHMT1/2 protect against pathological cardiac hypertrophy. *J Clin Invest* [Internet]. 2017 Jan 1 [cited 2023 Apr 22];127(1):335. Available from: [/pmc/articles/PMC5199699/](https://pmc/articles/PMC5199699/)
  - 46. Lee J, Kim K, Ryu TY, Jung CR, Lee MS, Lim JH, et al. EHMT1 knockdown induces apoptosis and cell cycle arrest in lung cancer cells by increasing CDKN1A expression. *Mol Oncol.* 2021 Nov 1;15(11):2989–3002.
  - 47. Rodríguez C, Martínez-González J. The Role of Lysyl Oxidase Enzymes in Cardiac Function and Remodeling. *Cells* [Internet]. 2019 Dec 1 [cited 2023 Apr 22];8(12). Available from: [/pmc/articles/PMC6953057/](https://pmc/articles/PMC6953057/)
  - 48. Wang TH, Hsia SM, Shieh TM. Lysyl Oxidase and the Tumor Microenvironment. *Int J Mol Sci* [Internet]. 2017 Jan 1 [cited 2023 Apr 22];18(1). Available from: [/pmc/articles/PMC5297697/](https://pmc/articles/PMC5297697/)

49. Peppino G, Ruiu R, Arigoni M, Riccardo F, Iacoviello A, Barutello G, et al. Teneurins: Role in Cancer and Potential Role as Diagnostic Biomarkers and Targets for Therapy. *Int J Mol Sci* [Internet]. 2021 Mar 1 [cited 2023 Apr 22];22(5):1–23. Available from: [/pmc/articles/PMC7956758/](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7956758/)
50. Zhu C, Feng X, Ye G, Huang T. Meta-analysis of possible role of cadherin gene methylation in evolution and prognosis of hepatocellular carcinoma with a PRISMA guideline. *Medicine (Baltimore)* [Internet]. 2017 [cited 2023 Apr 22];96(16). Available from: [/pmc/articles/PMC5406084/](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5406084/)
51. Gheldof A, Berx G. Cadherins and Epithelial-to-Mesenchymal Transition. *Prog Mol Biol Transl Sci.* 2013 Jan 1;116:317–36.
52. Krolevets M, Cate V ten, Prochaska JH, Schulz A, Rapp S, Tenzer S, et al. DNA methylation and cardiovascular disease in humans: a systematic review and database of known CpG methylation sites. *Clin Epigenetics* 2023 151 [Internet]. 2023 Mar 30 [cited 2023 Jun 10];15(1):1–16. Available from: <https://clinicalepigeneticsjournal.biomedcentral.com/articles/10.1186/s13148-023-01468-y>
53. Anglim PP, Galler JS, Koss MN, Hagen JA, Turla S, Campan M, et al. Identification of a panel of sensitive and specific DNA methylation markers for squamous cell lung cancer. *Mol Cancer* [Internet]. 2008 Jul 10 [cited 2023 Jun 10];7(1):1–13. Available from: <https://molecular-cancer.biomedcentral.com/articles/10.1186/1476-4598-7-62>
54. Butler M, Pongor L, Su YT, Xi L, Raffeld M, Quezado M, et al. MGMT status as a clinical biomarker in glioblastoma. *Trends in cancer* [Internet]. 2020 May 1 [cited 2023 Jun 10];6(5):380. Available from: [/pmc/articles/PMC7315323/](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7315323/)
55. Spitzwieser M, Holzweber E, Pfeiler G, Hacker S, Cichna-Markl M. Applicability of HIN-1, MGMT and RASSF1A promoter methylation as biomarkers for detecting field cancerization in breast cancer. *Breast Cancer Res* [Internet]. 2015 Sep 14 [cited 2023 Jun 10];17(1):1–13. Available from: <https://breast-cancer-research.biomedcentral.com/articles/10.1186/s13058-015-0637-5>
56. Ding Y, Yang Q, Wang B, Ye G, Tong X. The Correlation of MGMT Promoter Methylation and Clinicopathological Features in Gastric Cancer: A Systematic Review and Meta-Analysis. *PLoS One* [Internet]. 2016 Nov 1 [cited 2023 Jun 10];11(11). Available from: <https://pubmed.ncbi.nlm.nih.gov/27824946/>
57. Tan SX, Hu RC, Xia Q, Tan YL, Liu JJ, Gan GX, et al. The methylation profiles of PRDM promoters in non-small cell lung cancer. *Onco Targets Ther* [Internet]. 2018 May 22 [cited 2023 Jun 10];11:2991. Available from: [/pmc/articles/PMC5973400/](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5973400/)
58. Yamato G, Kawai T, Shiba N, Ikeda J, Hara Y, Ohki K, et al. Genome-wide DNA methylation analysis in pediatric acute myeloid leukemia. *Blood Adv.* 2022 Jun 14;6(11):3207–19.
59. Cuttini E, Goi C, Pellarin E, Vida R, Brancolini C. HDAC4 in cancer: A multitasking platform to drive not only epigenetic modifications. *Front Mol Biosci.* 2023 Jan 24;10:1116660.
60. Shi Y, Zhang H, Huang S, Yin L, Wang F, Luo P, et al. Epigenetic regulation in cardiovascular disease: mechanisms and advances in clinical trials. *Signal Transduct Target Ther* 2022 71 [Internet]. 2022 Jun 25 [cited 2023 Jun 15];7(1):1–28. Available from: <https://www.nature.com/articles/s41392-022-01055-2>
61. Zheng X, Gai X, Ding F, Lu Z, Tu K, Yao Y, et al. Histone acetyltransferase PCAF Up-regulated cell apoptosis in hepatocellular carcinoma via acetylating histone H4 and

- inactivating AKT signaling. Mol Cancer [Internet]. 2013 Aug 27 [cited 2023 Jun 15];12(1):1–11. Available from: <https://molecular-cancer.biomedcentral.com/articles/10.1186/1476-4598-12-96>
62. Kim K, Ryu TY, Jung E, Han TS, Lee J, Kim SK, et al. Epigenetic regulation of SMAD3 by histone methyltransferase SMYD2 promotes lung cancer metastasis. Exp Mol Med 2023 555 [Internet]. 2023 May 1 [cited 2023 Jun 15];55(5):952–64. Available from: <https://www.nature.com/articles/s12276-023-00987-1>
63. Anderson RA, Roussel AM, Zouari N, Mahjoub S, Matheau JM, Kerkeni A. Potential Antioxidant Effects of Zinc and Chromium Supplementation in People with Type 2 Diabetes Mellitus. <http://dx.doi.org/101080/07315724200110719034> [Internet]. 2013 [cited 2023 Jun 9];20(3):212–8. Available from: <https://www.tandfonline.com/doi/abs/10.1080/07315724.2001.10719034>
64. Domingo-Relloso A, Grau-Perez M, Galan-Chilet I, Garrido-Martinez MJ, Tormos C, Navas-Acien A, et al. Urinary metals and metal mixtures and oxidative stress biomarkers in an adult population from Spain: The Hortega Study. Environ Int [Internet]. 2019 Feb [cited 2019 Jul 8];123:171–80. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/30529889>
65. Little PJ, Bhattacharya R, Moreyra AE, Korichneva IL. Zinc and cardiovascular disease. Nutrition [Internet]. 2010 Nov 1 [cited 2018 May 22];26(11–12):1050–7. Available from: <https://www.sciencedirect.com/science/article/pii/S0899900710001036>
66. Grau-Perez M, Voruganti VS, Balakrishnan P, Haack K, Goessler W, Franceschini N, et al. Genetic variation and urine cadmium levels: ABCC1 effects in the Strong Heart Family Study. Environ Pollut [Internet]. 2021 May 5 [cited 2023 Jun 15];276:116717. Available from: [/pmc/articles/PMC8026674/](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8026674/)
67. Tellez-Plaza M, Guallar E, Howard B V., Umans JG, Francesconi KA, Goessler W, et al. Cadmium Exposure and Incident Cardiovascular Disease. Epidemiology [Internet]. 2013 May [cited 2023 Jun 9];24(3):421. Available from: [/pmc/articles/PMC4142588/](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4142588/)
68. Tellez-Plaza M, Jones MR, Dominguez-Lucas A, Guallar E, Navas-Acien A. Cadmium Exposure and Clinical Cardiovascular Disease: a Systematic Review. Curr Atheroscler Rep [Internet]. 2013 Oct 1 [cited 2023 Jun 9];15(10). Available from: [/pmc/articles/PMC3858820/](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3858820/)
69. Hartwig A. Cadmium and cancer. Met Ions Life Sci [Internet]. 2013 [cited 2023 Jun 9];11:491–507. Available from: <https://pubmed.ncbi.nlm.nih.gov/23430782/>
70. García-Esquinas E, Pollan M, Tellez-Plaza M, Francesconi KA, Goessler W, Guallar E, et al. Cadmium exposure and cancer mortality in a prospective cohort: the strong heart study. Environ Health Perspect [Internet]. 2014 [cited 2023 Jun 9];122(4):363–70. Available from: <https://pubmed.ncbi.nlm.nih.gov/24531129/>
71. Lamas GA, Bhatnagar A, Jones MR, Mann KK, Nasir K, Tellez-Plaza M, et al. Contaminant Metals as Cardiovascular Risk Factors: A Scientific Statement From the American Heart Association. J Am Heart Assoc [Internet]. 2023 Feb 14 [cited 2023 Jun 19];12:29852. Available from: <https://www.ahajournals.org/doi/abs/10.1161/JAHA.123.029852>

**Table 1. Participant characteristics**

	Non cases	Cancer cases	CVD cases	Cancer-CVD cases
<b>Strong Heart Study</b>	N=944	N=277	N=823	N=142
Age	52 (48, 59)	56 (50, 63)	56 (50, 63)	56 (50, 63)
Female %	61.1	54.2	56.5	49.3
Current smoking %	34.7	46.3	40.1	47.2
BMI	29.2 (25.6, 33.1)	29.2 (25.6, 33.9)	30.1 (27.1, 34.2)	30.0 (27.3, 34.5)
Diabetes %	32.8	37.7	52.6	45.8
LDL-cholesterol, mg/dL	115 (94, 137)	118 (101, 138)	123 (102, 144)	120 (103, 140)
HDL-cholesterol, mg/dL	45 (38, 54)	43 (37, 53)	42 (36, 50)	42 (36, 49)
Hypertension %	14.5	16.5	27.1	21.1
SBP, mmHg	121 (111, 132)	124 (111, 136)	127 (116, 140)	125 (111, 138)
<b>Framingham Heart Study</b>	N=864	N=337	N=222	N=51
Age	65 (59, 74)	66 (61, 74)	73 (65, 79)	75 (67, 80)
Female %	39.9	45.3	48.9	38.8
Current smoking %	50.5	56.5	56.2	55.1
BMI	30.5 (25.6, 98)	31.1 (25.8, 101)	35.3 (28.4, 127.5)	47.3 (29.6, 137.0)
Diabetes %	8.4	9.3	16	16.3
LDL-cholesterol, mg/dL	75 (0, 112)	80 (0, 113)	59 (0, 99.5)	20 (0, 93)
HDL-cholesterol, mg/dL	59 (43, 72)	60 (43.5, 71)	51 (41, 68)	57 (43, 71)
Hypertension %	21.6	21.6	26.5	32.7
SBP, mmHg	122 (99, 142)	126 (105, 154)	131 (101.5, 155)	121 (85, 143)
<b>Atherosclerosis Risk in Communities (African American)</b>	N=957	N=389	N=622	N=244
Age	54 (51, 59)	56 (51, 61)	57 (52, 62)	58 (53, 62)
Female %	71.7	49.4	62.4	51.2
Current smoking %	19.6	29.3	24.8	35.7
BMI	28.6 (25.7, 32.5)	28.6 (25.3, 32.5)	29.8 (26.6, 34.0)	29.6 (25.7, 33.7)
Diabetes %	12.7	15.9	32.0	23.0
LDL-cholesterol, mg/dL	130.8 (106.1, 155.1)	129.9 (106.4, 157.2)	135.9 (109.0, 162.6)	132.9 (108.2, 158.4)
HDL-cholesterol, mg/dL	16 (11, 21)	14 (10, 19)	13 (10, 19)	13 (10, 17)
Hypertension %	37.8	35.2	57.1	54.9
SBP, mmHg	120 (110, 132)	121 (110, 132)	128 (116, 143)	129 (119, 141)
<b>Atherosclerosis Risk in Communities (European American)</b>	N=404	N=193	N=222	N=84

Age	58 (54, 63)	59 (56, 64)	61 (57, 65)	61 (58, 64)
Female %	68.3	51.8	51.8	41.7
Current smoking %	15.1	19.7	22.5	22.6
BMI	25.3 (22.8, 27.7)	25.7 (23.4, 28.4)	26.3 (23.6, 29.5)	26.5 (24.5, 28.2)
Diabetes %	3.0	5.2	7.2	3.6
LDL-cholesterol, mg/dL	126.1 (106.2, 147.8)	127.2 (107.8, 145.3)	129.4 (107.0, 154.4)	127.2 (108.2, 145.8)
HDL-cholesterol, mg/dL	13 (9, 20)	12 (9, 18)	12 (7, 18)	12 (9, 16)
Hypertension %	13.6	14.5	21.6	14.3
SBP, mmHg	111 (104, 122)	112 (105, 124)	118 (110, 133)	115 (107, 129)

Median (interquartile range) and percentages are shown for continuous and categorical variables, respectively.

**Table 2. Meta-analyzed hazard ratios (95 % confidence intervals) for the common DMPs in the multi-cohort analysis for CVD, cancer, and cancer-CVD.**

CpG	Gene	Effect estimate	P-value	I2 statistic (%)
<b>CVD</b>				
cg06630010	<i>WDR37</i>	-1.62	0.0015	32.2
cg19693031	<i>TXNIP</i>	-1.14	0.0019	0
cg27153400	<i>ISOC2</i>	1.21	0.0033	0
cg14978242	<i>SERINC5</i>	-3.21	0.0036	60.7
cg07444323	<i>ADAM33</i>	1.25	0.010	45.5
cg16391101	<i>EHMT1</i>	8.07	0.010	37.1
cg01414728	<i>LINC01257</i>	-1.57	0.034	61.5
cg15111469	<i>LOX</i>	0.69	0.092	76.2
cg25957599	<i>TBXT</i>	4.39	0.093	68.5
<b>Cancer</b>				
cg11091196	<i>TENM2</i>	-2.13	0.00017	0
<b>Cancer-CVD</b>				
cg12107018	<i>CNPY1</i>	4.97	0.0019	53.2
cg08449049	<i>CDH15</i>	1.54	0.028	66.9

All models were adjusted for age, sex, BMI, smoking status (never, former, current), DNA methylation-based smoking score and estimated cell counts. CVD and Ca-CVD models were additionally adjusted for LDL cholesterol, HDL cholesterol, diabetes (yes/no), hypertension treatment (yes/no) and systolic blood pressure. SHS models were additionally adjusted for study center (Arizona, Oklahoma or North Dakota / South Dakota), five genetic PCs that accounted for population stratification and, for CVD and Ca-CVD models, albuminuria (microalbuminuria, normal albumin levels or macroalbuminuria). FHS models were additionally adjusted for batch, as DNAm data were processed in two different batches. ARIC models additionally adjusted for field center (Jackson, MS; Washington County, MD; suburban Minneapolis, MN; and Forsyth County, NC), and five PCs to correct batch effects in each cohort.

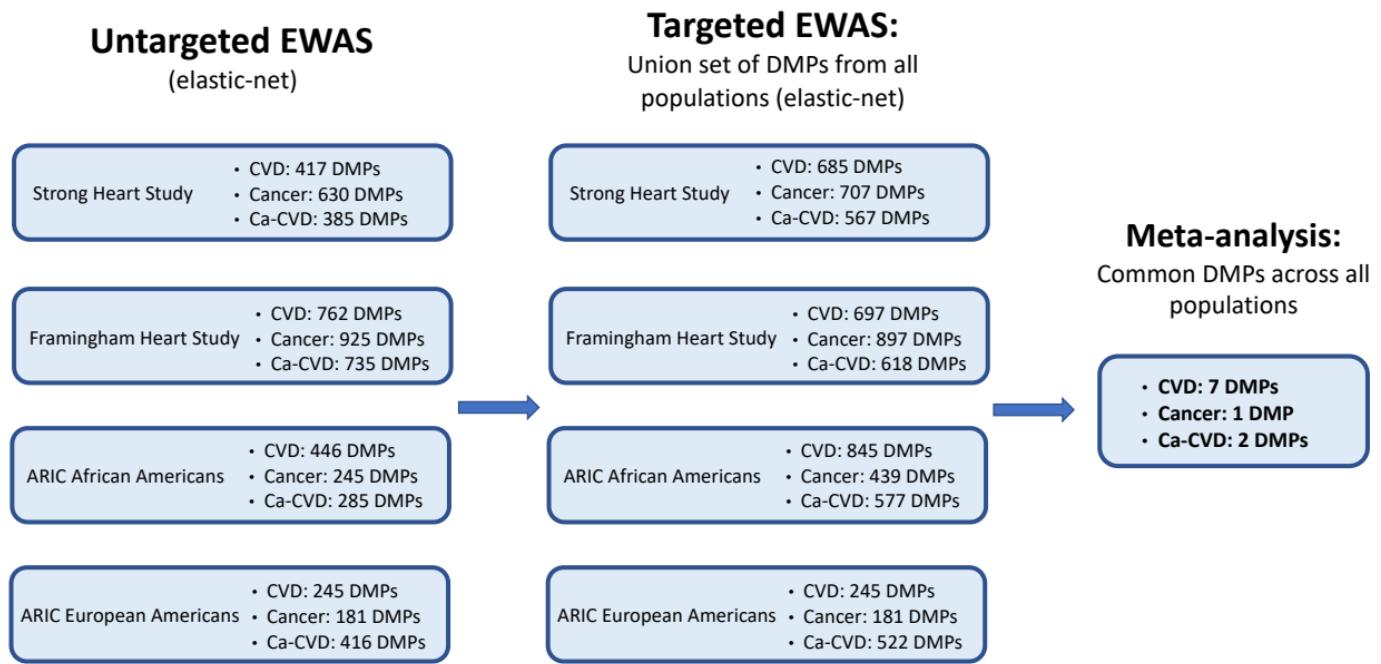
**Table 3. Genes classified as drug targets in the DrugBank database.** The 69 common nodes/genes in overlapping Ca, CVD and Ca-CVD DMPs in the multi-cohort networks were searched.

Drug ID	Drug name	Description and/or indication	Target (UniProt ID)
DB02216	S-Methylcysteine	This compound belongs to the class of organic compounds known as l-cysteine-s-conjugates. These are compounds containing L-cysteine where the thio-group is conjugated	MGMT (P16455)
DB04531	S-Benzylcysteine	This compound belongs to the class of organic compounds known as l-cysteine-s-conjugates. These are compounds containing L-cysteine where the thio-group is conjugated	
DB00151	Cysteine	Cysteine is an amino acid commonly found as a component of total parenteral nutrition and used as an antidote for acetaminophen overdose.	
DB01593	Zinc	Zinc is an essential element commonly used for the treatment of patients with documented zinc deficiency.	
DB11831	Dinitrochlorobenzene	Dinitrochlorobenzene has been used in trials studying the treatment of HIV Infections.	
DB14487	Zinc acetate	Zinc acetate is a medication used to treat zinc deficiency.	
DB14533	Zinc chloride	Zinc chloride injections are indicated for use total parenteral nutrition to maintain zinc serum levels and prevent deficiency syndromes.	
DB14548	Zinc sulfate, unspecified form	Zinc sulfate is a common zinc supplement in parenteral nutrition	PTK6 (Q13882)
DB05294	Vandetanib	Vandetanib is an oral once-daily kinase inhibitor of tumour angiogenesis and tumour cell proliferation with the potential for use in a broad range of tumour types.	
DB12010	Fostamatinib	Fostamatinib has been investigated for the treatment and basic science of Rheumatoid Arthritis and Immune Thrombocytopenic Purpura (ITP)	
DB15035	Zanubrutinib	Zanubrutinib is a kinase inhibitor used to treat mantle cell lymphoma, a type of B-cell non-Hodgkin lymphoma, in adults who previously received therapy.	
DB11800	Tivozanib	Tivozanib is a kinase inhibitor to treat adult patients with renal cell carcinoma (RCC) who have failed prior systemic therapies or experienced relapsed disease.	

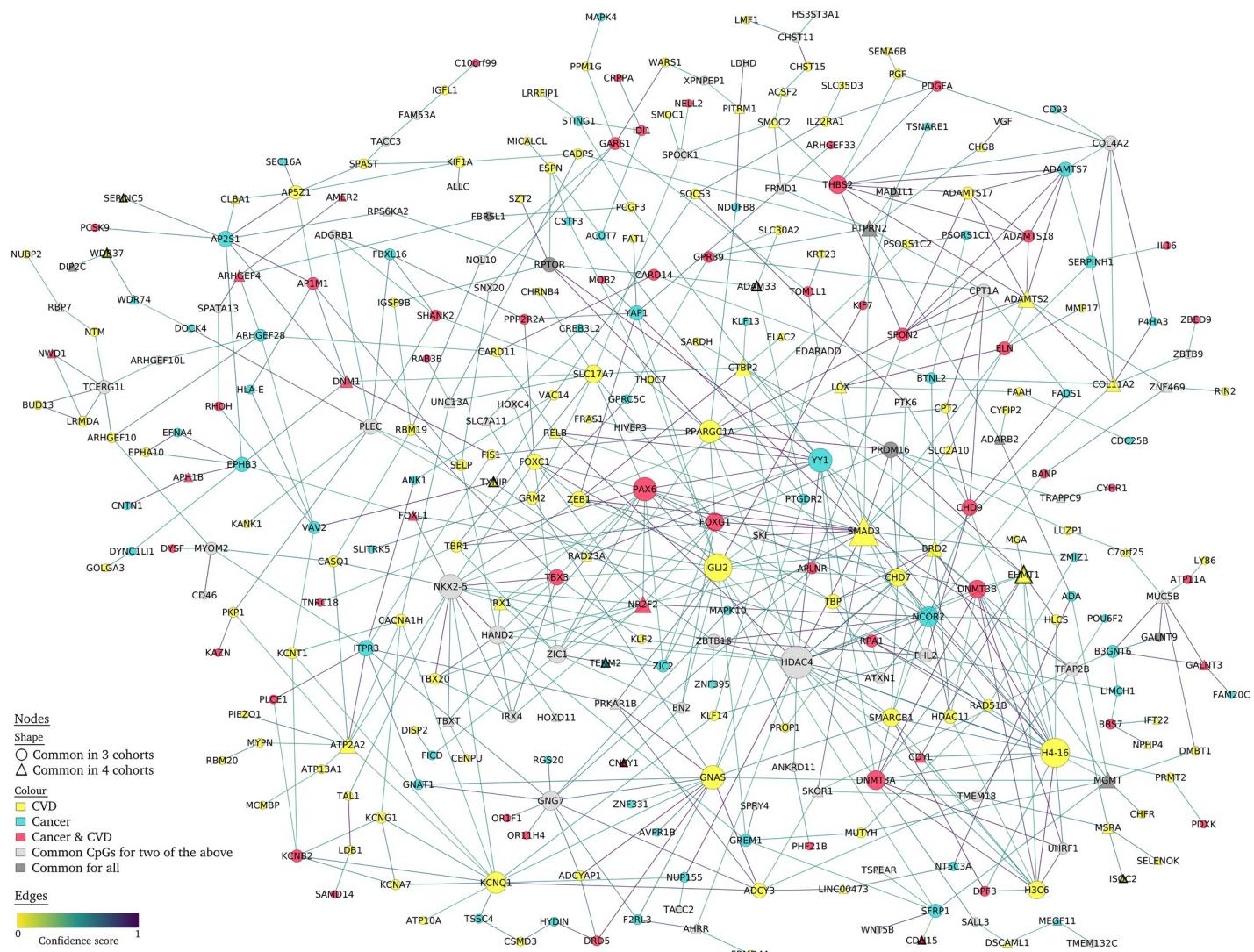
DB01812	Adenosine 3',5'-diphosphate	Not Available	HS3ST3A1 (Q9Y663)
DB02264	O2-Sulfo-Glucuronic Acid	Not Available	
DB03959	N,O6-Disulfo-Glucosamine	Not Available	
DB03981	1,4-Dideoxy-5-Dehydro-O2-Sulfo-Glucuronic Acid	Not Available	
DB04380	Allantoate	Not Available	ALLC (P77425)
DB00583	Levocarnitine	Levocarnitine is a quaternary ammonium compound used to treat carnitine deficiency or to stimulate gastric and pancreatic secretions in hyperlipoproteinemia.	CPT1A (P50416)
DB01074	Perhexiline	Perhexiline is a coronary vasodilator used especially for angina of effort. It may cause neuropathy and hepatitis.	
DB01016	Glyburide	Glyburide is a sulfonylurea used in the treatment of non insulin dependent diabetes mellitus.	
DB03309	N-cyclohexyltaurine	Not Available	EN2 (P19622)
DB07879	N-hydroxy-5-[(3-phenyl-5,6-dihydroimidazo[1,2-a]pyrazin-7(8H)-yl)carbonyl]thiophene-2-carboxamide	Not Available	HDAC4 (P56524)
DB08613	2,2,2-TRIFLUORO-1-{5-[(3-PHENYL-5,6-DIHYDROIMIDAZO[1,2-A]PYRAZIN-7(8H)-YL)CARBONYL]THIOPHEN-2-YL}ETHANE-1,1-DIOL	Not Available	
DB06176	Romidepsin	Romidepsin is a histone deacetylase (HDAC) inhibitor used to treat cutaneous T-cell lymphoma.	
DB01593	Zinc	Zinc is an essential element commonly used for the treatment of patients with documented zinc deficiency.	
DB14487	Zinc acetate	Zinc acetate is a medication used to treat zinc deficiency.	
DB14533	Zinc chloride	Zinc chloride is a medication used to treat zinc deficiencies and associated symptoms and also in total parenteral nutrition.	
DB14548	Zinc sulfate, unspecified form	Zinc sulfate, unspecified form is a zinc supplement indicated in parenteral nutrition.	
DB00162	Vitamin A	Vitamin A is a vitamin important for retinal function that is used clinically to correct vitamin A deficiency	RBP7 (Q96R05)
DB03940	Oxamic Acid	Not Available	LDHD (P30901)



**Figure 1.** Flowchart of the multi-cohort analysis.

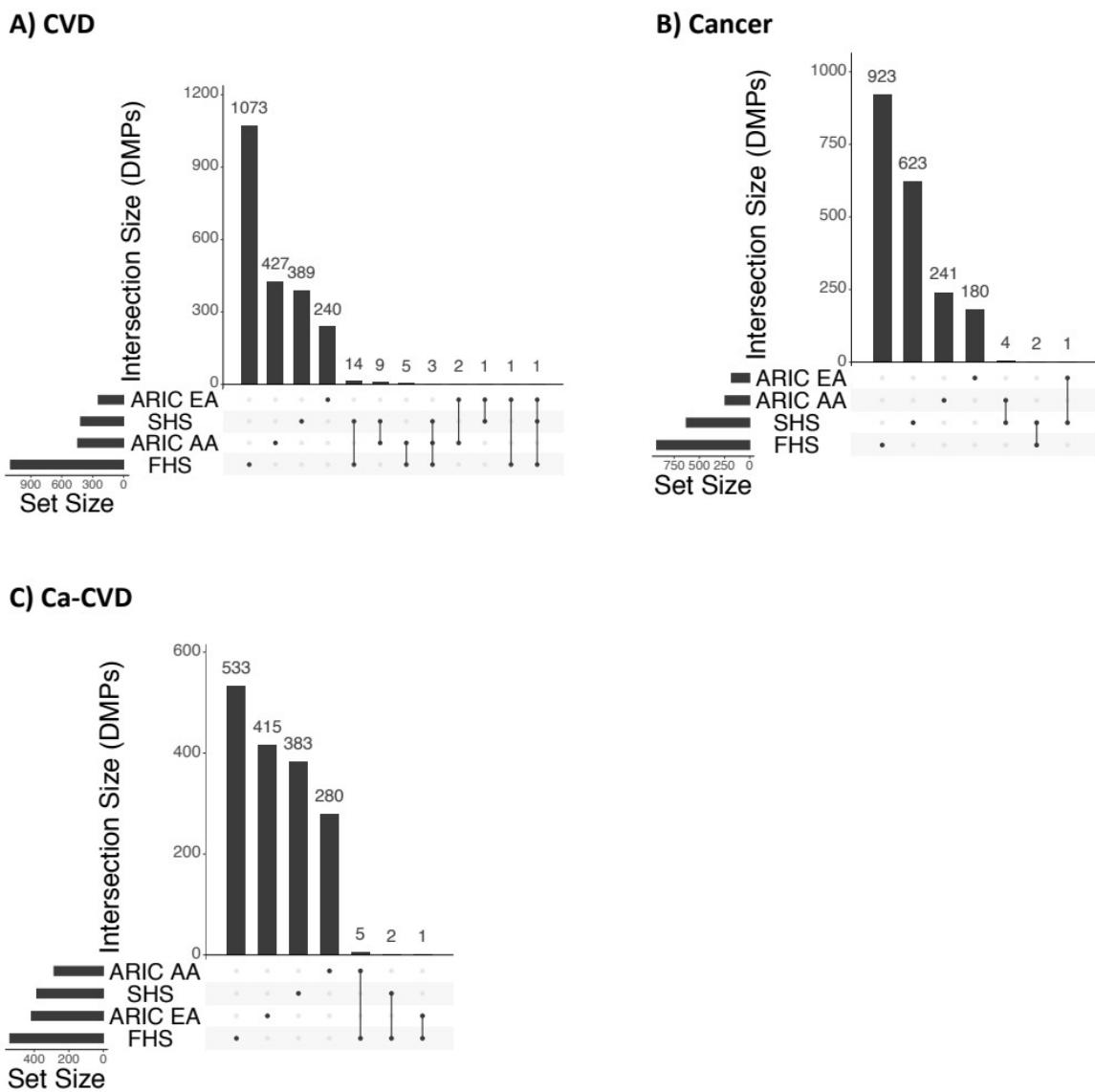


**Figure 2. Protein-protein interaction network for cancer and cardiovascular diseases for protein-coding genes associated with at least 3 cohorts for Ca, CVD and Ca-CVD endpoints and corresponding directly-related nodes.** The network is displayed by Cytoscape and contains 312 nodes and 574 interactions. The size of the nodes is proportional to the number of connections. Increasingly darker solid edge lines indicate protein interaction with increasing confidence scores. The interactions and its confidence score of 0.4 or greater, were obtained from STRING database. Thick-edged nodes indicate meta-analysis results.



## SUPPLEMENTARY MATERIAL

**Figure S1.** Upset plot of the intersection of DMPs across cohorts in the untargeted analysis for  
**a) CVD b) Cancer c) Cancer-CVD.**



**Table S1. Predictive ability of different sets of DMPs for cancer in individuals with previous CVD events.**

N CpGs	All Cause Cancer (N cases/non cases=127/823)	
Accumulated follow-up time, person-years	9185.2	
Reference: fully adjusted model without DMPs*	0	0.65
Reference + Ca DMPs	626	0.80
Reference + Ca-CVD DMPs	380	0.96
Reference + DMPs annotated to overlapping genes for any combination of Ca, CVD and Ca-CVD DMPs (green nodes in the protein interactions network)	265	0.83

\* Adjusted for age, sex, BMI, smoking status (never, former, current), DNA methylation-based smoking score, estimated cell counts, LDL cholesterol, HDL cholesterol, diabetes (yes/no), hypertension treatment (yes/no), systolic blood pressure, study center (Arizona, Oklahoma or North Dakota / South Dakota), five genetic PCs that accounted for population stratification and albuminuria (microalbuminuria, normal albumin levels or macroalbuminuria).

**Supplementary File 1. Excel file with all the detailed results obtained in the different models.** The excel sheets were named as A, B, C, D, E, F, G, H, I, J, K, L, M, N, O and P in the main text and correspond to the following results in the excel file:

- A) CVD DMPs in the SHS.** The CpG sites selected by the elastic-net model in primary endpoints analysis for CVD group
- B) Cancer DMPs in the SHS.** The CpG sites selected by the elastic-net model in primary endpoints analysis for Ca group
- C) Ca-CVD DMPs in the SHS.** The CpG sites selected by the elastic-net model in primary endpoints analysis for Ca-CVD group
- D) CVD DMPs in the FHS.** The CpG sites selected by the elastic-net model in primary endpoints analysis for CVD group
- E) Cancer DMPs in the FHS.** The CpG sites selected by the elastic-net model in primary endpoints analysis for Ca group
- F) Ca-CVD DMPs in the FHS.** The CpG sites selected by the elastic-net model in primary endpoints analysis for Ca-CVD group
- G) CVD DMPs in ARIC African Americans.** The CpG sites selected by the elastic-net model in primary endpoints analysis for CVD group
- H) Cancer DMPs in ARIC African Americans.** The CpG sites selected by the elastic-net model in primary endpoints analysis for Ca group
- I) Ca-CVD DMPs in ARIC African Americans.** The CpG sites selected by the elastic-net model in primary endpoints analysis for Ca-CVD group
- J) CVD DMPs in ARIC European Americans.** The CpG sites selected by the elastic-net model in primary endpoints analysis for CVD group
- K) Cancer DMPs in ARIC European Americans.** The CpG sites selected by the elastic-net model in primary endpoints analysis for Ca group
- L) Ca-CVD DMPs in ARIC European Americans.** The CpG sites selected by the elastic-net model in primary endpoints analysis for Ca-CVD group
- M) Network Nodes.** The 312 nodes included in the protein interaction network (Figure 2) and corresponding statistical parameters.
- N) Network Edges.** Confidence score and other edge parameters for each of the 574 interactions included in the network (Figure 2).
- O) Network enrichment analysis.** The specific subnetwork (nodes and interaction) included in the analyses and the main results.
- P) DrugBank database target search.** 69 overlapping nodes/genes in overlapping Ca, CVD and Ca-CVD DMPs. Accessed March 17, 2023.