Epigenome-wide association study of all-cause cancer and cardiovascular disease:

Common blood DNA methylation signatures in the Strong Heart Study

Angela L. Riffo-Campos[1], Arce Domingo-Relloso[2,3,4], Guillermo Ayala[4], Karin Haack[5], Andrés Ávila[1], Dorothy Rhoades[6], Jason G. Umans[7], M Daniele Fallin[7], Marina Pollan[2], Shelley Cole[5], Ana Navas-Acien[3] and Maria Tellez-Plaza[2]

[1]Centro de Excelencia de Modelación y Computación Científica, Universidad de La Frontera, Temuco, Chile.
[2]Department of Chronic Diseases Epidemiology, National Center for Epidemiology, Carlos III Health Institute, Madrid, Spain.
[3]Department of Environmental Health Sciences, Columbia University Mailman School of Public Health, New York, US.
[4]Departamento de Estadística e Investigación Operativa, Universidad de Valencia, Valencia, España.
[5] Population Health Program, Texas Biomedical Research Institute, San Antonio, TX, US.
[6]Stephenson Cancer Center, University of Oklahoma Health Sciences Center, Oklahoma City, Oklahoma, USA
[7] MedStar Health Research Institute, Washington DC, USA

***Co-corresponding authors:**

Angela Leticia Riffo-Campos
Centro de Excelencia de Modelación y Computación Científica
Universidad de La Frontera
Montevideo Street, 0785
4780000 Temuco-Chile
E-mail address: angela.riffo@ufrontera.cl

Ana Navas-Acien, MD, PhD
Professor of Environmental Health Sciences
Columbia University
Mailman School of Public Health
722 W 168th Street
Office 1105F
New York, NY 10032
an2737@cumc.columbia.edu

Running title: Common epigenetic signatures for cancer and cardiovascular diseases incidence

Abstract

Cancer (Ca) and cardiovascular diseases (CVD) are generally studied separately. Emerging evidence, however, reveals their complex inter-relationships. At the population level, we evaluated epigenetic signatures common for Ca and CVD and assessed their role in relevant biological pathways. Baseline DNA methylation (DNAm) profiles of 2186 participants from the Strong Heart Study (SHS) without prevalent Ca nor CVD and with complete information on potential confounders and follow-up were analyzed using the Infinium MethylationEPIC BeadChip. We classified incident cases of only Ca (N=277), only CVD (N=823), and both Ca and CVD (Ca-CVD) (N=142). Through a Cox proportional hazards model regularized by an elastic net penalty approach, we identified 178, 558, and 371 differentially methylated positions (DMPs) for Ca, CVD, and Ca-CVD, respectively. The enrichment analysis of genes attributed to identified DMPs largely pointed to known biological pathways involved in Ca and CVD. Identification of potential treatment targets in the DrugBank database among genes attributed to the overlap of Ca, CVD and Ca-CVD DMPs pointed to essential elements as potential compounds targeting the corresponding gene products. In an additional analysis among 935 participants with an incident CVD event (112 of them with a subsequent incident Ca), incident Ca predictive accuracy substantially increased, comparing models without and with the 371 Ca-CVD DMPs. Our results suggest that screening based on common epigenetic signatures of Ca and CVD among newly diagnosed CVD patients may identify patients at increased Ca risk, and potentially enable the precision prevention and control of Ca.

Key words: Cardio-oncology, DNA methylation, American Indians.

Introduction

Cardiovascular disease (CVD) and cancer are the most common causes of death worldwide. Emerging evidence reveals the complex relationship between cancer and CVD (1). Both cancer and CVD have risk factors in common, including smoking, physical inactivity, unhealthy diet, excess adiposity, genetic predisposition, environmental toxicants, advanced age and others (2–5). Cardiotoxicity, in addition, is an emerging problem following cancer treatment, such as chemotherapy, radiotherapy (4–7), and, to a lesser extent, also the targeted cancer immunotherapies (8). Both diseases share several biological pathways, including chronic inflammation, oxidative stress, aberrant apoptosis, and angiogenesis (3,4,9). A deeper understanding of the molecular relationships between these two diseases can contribute to identify new and safer therapeutic approaches for cancer and CVD.

Epigenetic modifications refer to the study of heritable changes that have effects on gene regulation without changing the DNA sequence; they are involved in expression and replication, among others (10). Several epigenetic marks are associated with biological functions in health and disease (11). The most broadly studied epigenetic mark in epidemiologic studies is DNA methylation (DNAm), which refers to the addition of a methyl group in the carbon-5 position of cytosine followed by guanine separated by a phosphate (CpG) (11). Traditionally, cancer and CVD epigenetics changes have been studied separately (12,13). Although there is increasing evidence supporting the relationship between CVD and cancer, the joint association of DNAm with cardiovascular and cancer incidence has not been evaluated.

In this study, we aimed to identify common epigenetic modifications for CVD and cancer by studying the overlap of differentially methylated positions (DMPs) for cardiovascular (CVD), cancer (Ca), and both (Ca-CVD) from separate epigenome-wide association studies. The study was conducted in the Strong Heart Study, the oldest and largest study of CVD and its risk factors in American Indian communities from the Northern Plains, the Southern Plains, and the Southwest. We used a Cox proportional hazards model regularized by an elastic net penalty approach, to jointly

model a panel of ~850000 CpGs measured by the Infinium MethylationEPIC BeadChip. We subsequently conducted molecular pathways analyses to assess potential downstream biological implications of DMPs associated to both Cancer and CVDs (i.e. integrative analysis of protein-to-protein interaction networks among proteins encoded by genes annotated to identified CVD-, Ca- and Ca-CVD-DMPs, enrichment analyses using GO terms, and identification of drug targets from DrugBank database). Finally, to explore if DNA methylation can help with early detection of cancer in patients who had a previous cardiovascular disease event, we evaluated the predictive ability of previously identified DMP sets to identify a future Ca event following a newly diagnosed cardiovascular event.

Methods

Study population. The SHS is a prospective cohort study funded by the National Heart, Lung and Blood Institute to investigate cardiovascular diseases 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 3517 potential participants. We further excluded 252 participants with prevalent cardiovascular disease, 136 participants with prevalent cancer, 429 participants missing urinary metals (as the blood DNA analysis was funded in the context of a study on metals and cardiovascular risk), 44 participants missing data on other risk factors, and 469 lacking sufficient DNA samples. After DNAm analysis, we removed data from 18 participants without classical DNAm bimodal distribution and data from 8 individuals with low median intensity levels, leaving 2186 participants for analyses.

Microarray DNA methylation determinations. 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 and DNAm was measured using the Illumina's MethylationEPIC BeadChip, which provides a measure of DNAm at a single nucleotide resolution at >850,000 CpG sites. This platform is enriched with DNA regions related to disease and likely labile to environmental exposures. 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. Detection of failed probes was carried out according to Illumina's recommendations.

We conducted sample quality control based on Illumina 850K control probes to assess staining, hybridization, bisulfite conversion, and other parameters (15). We assessed the correlation of replicate samples across plates, enabling us to identify any potential problems with particular plates/batches and to assess the accuracy of the DNAm values reported. We determined the total intensity (methylated + unmethylated channel) across all probes measured and any samples with extremely low overall intensity were removed. Finally, we ran single sample noob normalization as conducted by R package minfi (15–17). This is a background correction method with dye-bias normalization for Illumina Infinium methylation arrays. Principal Component analysis was done to determine if any batch effect correction was required. Blood cell count heterogeneity and batch effect adjustment was conducted using R package sva (18). We annotated CpGs to the nearest gene (reference genome hg19) using the 'matchGenes' function in 'minfi' R package (16).

Cancer and CVD incidence ascertainment and follow-up definition. The SHS uses tribal records, death certificates, medical records, and direct annual contact with participants and their families to assess health outcomes and vital status over time to identify fatal and non-fatal CVD and Ca, which were assessed by annual mortality and morbidity surveillance reviews of hospitalization and death records through 2017 and at two research clinic visits conducted in 1993–1995 and 1998–

1999. Medical records were reviewed by mortality and morbidity review committees composed of physician reviewers who assigned cancer and cardiovascular events.

We grouped the primary endpoints in three groups: all-cause incident cancer (Ca) in participants that did not develop incident CVD, all-cause incident CVD in participants that did not develop incident Ca, and combined cancer and CVD (Ca-CVD) for participants who developed both incident CVD and Ca during the follow-up (in either order). In a secondary analysis, we analyzed cause-specific Ca endpoints including lung, breast, colorectal, lymphatic-hematopoietic, esophageal-stomach, and CVD specific endpoints including coronary heart disease, stroke and heart failure. Detailed definitions of cause-specific fatal and nonfatal CVD events have been described elsewhere (14,19). For cancer-specific analyses, we only excluded prevalent cases of each cancer type.

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 Ca-CVD 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.

Other variables. Baseline sociodemographic, life-style and anthropometric information were obtained through interview and physical examination. The standardized in-person questionnaire included sociodemographic data (age, sex, education, study region), and smoking history (smoking status and pack-years), and body mass index (kg/m^2), diabetes, and HDL and LDL - cholesterol concentrations information were obtained during the Strong Heart Study questionnaires, collection of biological samples, laboratory determinations and physical exams.

Statistical Methods

Differentially methylated positions (DMP). We computed DNA methylation proportions at each CpG site (so-called "beta-values") and corresponding logit2 of DNAm methylation proportions (so-called "M-values"). Standard Cox regression cannot accommodate high-dimensional data or highly correlated predictors. Thus, we used GLMnet penalized Cox regression (Cox elastic-net, R package glmnet) (20,21) to simultaneously introduce all CpGs sites present in the MethylationEPIC BeadChip in the same model. The elastic-net framework is a mix between Ridge and Lasso regression (21). The algorithm fits a Cox model with penalty controlled by the α parameter, which can range from 0 –corresponding to Ridge regression, which can introduce more than one predictor from a correlated set, to 1 –corresponding to Lasso regression, which generally selects only one of the correlated predictors. We selected α=0.05, which is a popular elastic-net choice that works well for methylation data (22). The regularization parameter λ was selected using 10-folds cross-validation in our study. We fitted a Cox elastic-net model using Ca, CVD and Ca-CVD time-to-event outcomes in separate models to identify Ca-DMPs, CVD-DMPs and Ca-CVD-DMPs, respectively. All the CpG sites in the array were simultaneously included as predictors, including adjustment for age, sex, BMI, smoking status, study center, cell counts and five genetic PCs. CVD and Ca-CVD models were additionally adjusted for LDL cholesterol, HDL cholesterol, diabetes, hypertension treatment, systolic blood pressure and albuminuria.

Molecular pathways analyses. To explore molecular pathways associated to the identified DMPs, we conducted GO terms enrichment analysis of genes annotated to Ca-, CVD- and Ca-CVD-DMPs (see Figure 1), using missMethyl Bioconductor package (23). The information of health endpoints associated to the genes annotated to the identified DMPs was obtained from the Comparative Toxicogenomics Database (24), by using the GO term as input and sorted by number of references. In addition, the CpGs associated to protein coding genes were included in a protein-protein interaction network, after obtaining the interactions between the nodes from the STRING database v11.0 (25). All interaction sources with a confidence score of 0.1 or greater were included. The

confidence score (from 0 to 1) is provided by STRING database and estimate the likelihood that the annotated interaction between a given pair of proteins is biologically meaningful, specific and reproducible (see Szklarczyk et al. 2019 for a complete description). The network was analyzed and displayed using the yFiles organic layout with Cytoscape v3.7.1 (26). Further, a network enrichment analysis was performed as conducted by STRING using a subnetwork with the proteins in common for Ca and CVDs (i.e. proteins annotated to any combination of overlapping Ca, CVD and Ca-CVD DMPs, see green nodes in Figure 2) and corresponding directly connected nodes. The databases GO, KEGG, INTERPRO, UniProt and Reactome were included in the network enrichment analysis. The network enrichment analysis allows to incorporate the relationships between nodes within the gene enrichment analysis, using a combination of methods that include multiple testing correction, two-sided Kolmogorov–Smirnov test and hierarchical clustering of the STRING network itself (see Szklarczyk et al. 2019 for a complete description). In addition, we attempted to identify additional biological mechanisms potentially in common for Ca and CVD by evaluating which nodes are drug targets within DrugBank database v5.1.7. The DrugBank is a richly annotated database that combines detailed drug data with comprehensive drug target information (27) to assess if some of the genes reported in our results have been identified as drug targets, providing the corresponding targeting compound. Specifically, we searched among 5,225 non-redundant proteins linked to the drug entries in DrugBank and contrasted genes in common for Ca, CVD and Ca-CVD with those that were classified as drug targets.

Predictive models of incident cancer in an analysis restricted to participants who developed CVD first. To explore if DNAm might help with the early detection of cancer in patients with newly diagnosed CVD, we evaluated the predictive ability of baseline blood DNAm signatures, as measured by the concordance index (C-statistic), for cancer prediction and comparing different sets of candidate DMPs (i.e. combinations of identified Ca, CVD and Ca-CVD DMPs, and relevant

protein network nodes) in progressively adjusted models. Specifically, we restricted the analysis to participants who had a cardiovascular event at any time and excluded individuals that developed cancer 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). We conducted four progressively adjusted models: 1) only risk factors (no CpGs) (Model 1); 2) Model 1 further adjusted for the 178 DMPs identified in the Ca main models; 3) Model 1 further adjusted for the 383 DMPs identified in the Ca-CVD models; 4) Model 1 further adjusted for the 80 DMPs annotated to protein nodes from any combination of overlapping Ca, CVD and Ca-CVD DMPs (see Table S1).

Results

Incident Ca and CVD events. Over a ~28-year maximum follow-up, we observed 277, 823, and 142 Ca, CVD, and Ca-CVD cases, respectively (Table 1, Figure 1). The corresponding accumulated follow-up was 54,131, 34,301 and 20,650 person-years for Ca, CVD and Ca-CVD analysis, respectively. Among participants with Ca-CVD, 112 were diagnosed with CVD first (mean 6.4 years before the Ca diagnosis), 15 were diagnosed with Ca first (mean 4 years before the CVD diagnosis), and 15 were diagnosed with both diseases at the same time. Participants with incident CVD were older and more likely to have baseline albuminuria, hypertension, and diabetes compared to non-cases. Baseline current smoking was more common in participants with incident Ca, CVD and Ca-CVD, especially in those with both diseases (Table 1), compared to non-cases. CVD-specific endpoints included 647, 298, and 313 coronary heart disease (CHD), stroke and heart failure (HF) cases, respectively. Ca-specific endpoints included 73, 39, 32, 29, and 12, lung, breast, colorectal, lymphatic-hematopoietic, and esophageal/stomach Ca cases.

Differentially Methylated Positions. The elastic-net model selected 628, 178 and 383 DMPs for incident Ca, CVD and Ca-CVD, respectively (Supplementary File 1, sheets A, B and C). The C

indexes including only traditional risk factors and further including methylation data were, respectively, 0.63 and 0.76 for Ca, 0.73 and 0.81 for CVD, and 0.80 and 0.91 for Ca-CVD. We estimated hazard ratios and 95% confidence intervals for the CpGs selected by the elastic-net model using Cox proportional hazards models. At a nominal p-value of 0.05, 558 (out of the 628 selected), 178 (out of 178 selected) and 371 (out of 383 selected) were significant for CVD, Ca, and Ca-CVD respectively (Supplementary file 1, sheets A, B and C). The top ten Ca-DMPs were annotated to *TMEM132C*, *LGMN*, *GINS3*, *EPHB3*, *TBX18*, *SERPINH1*, *ALLC*, *WNT3A*, *TBC1D30* and *POM121L2* (Table 2). The top ten CVD-DMPs were annotated to *EXOSC10*, *LOC646522*, *POLE*, *EML6*, *EXOSC10*, *FMNL3*, *ACBD3*, *TRIM26*, *MYPN* and *PCP4L1* (Table 3). The top ten Ca-CVD-DMPs were mostly annotated to killer cell immunoglobulin-like receptors genes (Table 4). The CpG cg14391737 in the *PRSS23* gene was a DMP selected by the elastic-net models for the three endpoints (Ca, CVD and Ca-CVD) (Table 5).

The elastic-net model for specific CVD events selected 479 CpGs for CHD, 657 for HF and 586 for stroke (Supplementary File 1, sheets D, E and F); for specific Ca events it selected 622 CpGs for lung cancer, 156 for lymphatic-hematopoietic cancer, 8 for esophageal/stomach cancer, one for colorectal cancer and 11 for breast cancer (Supplementary File 1, sheets G, H, I, J and K). There were 6 common CpGs for CHD and lung cancer (cg24023859, cg04781566, cg21323642, cg10217853, cg10403212 and cg14391737), 4 common CpGs for heart failure and lung cancer (cg03648784, cg02440512, cg10217853 and cg08350488), and 1 common CpG for stroke with lymphatic-hematopoietic cancer (cg11223747), lung cancer (cg25544931), and esophageal/stomach cancer (cg26142075).

Molecular pathways analyses. A total of 294, 166 and 284 GO terms were significantly enriched for CVD, Ca and Ca-CVD analyses, respectively (Figure 1, supplementary file 1, sheets L, M and N). The top ten GO terms in CVD were related to hypertension, myocardial infarction (GO:0005184, GO:0018216, GO:0021894), cardiofaciocutaneous syndrome (GO:1903358), kidney

disease (GO:0042813), cancer (GO:0048387, GO:1990909, GO:0043208, GO:0043679) and arginine methyltransferases (GO:0035241). The top ten GO terms in Ca were related to hypertension (GO:0043400, GO:0044508, GO:0004967), cancer (GO:0045944, GO:0007263, GO:0015552, GO:0015730), diabetes (GO:0035883), stomach ulcer (GO:0031768) and liver injury (GO:0060730). The top ten GO terms in Ca-CVD were related to diabetes (GO:0050692, GO:0005899), cancer (GO:0008270, GO:0030539, GO:0043548, GO:0031017), myocardial infarction (GO:0045087), atherosclerosis (GO:0042277), liver injury (GO:0090461) and schizophrenia (GO:0070554). Among significant GO terms, 18 were common for the 3 groups (Table S2).

In protein-protein interaction network analysis, of the 1092 unique DMPs associated to Ca, CVD and Ca-CVD (178, 558 and 371 DMPs respectively, 15 of them were repeated between the 3 groups), we excluded 100 duplicate genes (Figure 1). Additionally, 276 ncRNA or protein-coding genes without connections in the network were discarded. As a result, the network included 716 nodes and 1622 edges (Figure 2, Supplementary File 1, sheets O and P). The most connected nodes were *EGFR* (associated to CVD DMPs, with 39 connections), *WNT3A* and *NKX2-5* (associated to Ca and Ca-CVD DMPs, with 39 and 35 connections, respectively), *TBX3* (associated to Ca, CVD and Ca-CVD DMPs, with 35 interactions) and *MYC* (associated to Ca-CVD DMPs, 32 interactions). These nodes were directly connected to *NEUROD1* and *HDAC4*, the most connected nodes in Ca-DMPs (28 and 24 interactions, respectively). The 39 overlapping nodes between at least two of Ca, CVD and Ca-CVD-related DMPs and their corresponding 199 directly connected-nodes were included in a network enrichment analysis. The results included Wnt signaling pathway (hsa04310, HSA-195721, KW-0879, HSA-3238698, FDR < 0.05), transcription regulation (KW-0805, GO:0045893, GO:0044212, FDR < 0.01), Hippo signaling pathway (hsa04390, FDR = 0.0005), AMPK signaling pathway (hsa04152, FDR = 0.014), p53-like transcription factor (IPR008967, FDR = 0.017), PI3K-Akt signaling pathway (hsa04151, FDR = 0.017), adherence junction (hsa04520, FDR = 0.012) and others (Supplementary File 1, Q sheet). Additionally, the 39

overlapping nodes for cancer and CVD -related endpoints were included in the DrugBank database. The JUP (P14923), TP73 (O15350), ANO1 (Q5XXA6), HLCS (P50747) and PCK2 (Q16822) proteins were identified as drug targets in the DrugBank database (Table 6).

Incident cancer prediction among individuals with a prior CVD event. 127 individuals developed cancer after an initial CVD event (accumulated follow-up time was 14131.2 person-years). The corresponding number of cancer specific endpoints (accumulated follow-up time) was 23 (15170.8 person-years) for lung, 15 (8318.5 person-years) for breast, 12 (15183 person-years) for lymphatic/hematopoietic and 11 (15169.6 person-years) for easophageal/stomach. The C-statistic for the fully adjusted model without blood DNAm ranged from 0.50 for esophageal/stomach cancer to 0.85 to lung cancer (Table S1). While the best predictive ability for a cancer event following a CVD event was observed when including the 383 Ca-CVD DMPs (C indexes ranged from 0.96 for all-causes and lymphatic/hematopoietic cancers to 0.99 for lung, breast and esophageal/stomach cancer), with only the 80 CpGs corresponding to the DMPs annotated to protein nodes from any combination of overlapping Ca, CVD and Ca-CVD DMPs (i.e. green nodes in Figure 2) the improvement in C-statistic was almost the same for cancer-specific endpoints.

DISCUSSION

Data from SHS adult participants recruited in 1989-91 and followed through 2017 has identified novel epigenetic signatures for all-cause Ca, CVD and Ca-CVD DMPs, including a total of 14 overlapping DMPs that were independently selected by at least 2 of the 3 models for each endpoint. At the gene level, a total of 43 overlapping genes were found. The proteins encoded by these genes and those directly connected with them, were associated to common molecular pathways for Ca and CVD. Consistently, in protein network enrichment analysis, nodes related to novel as well as to widely known mechanisms for both diseases were interconnected, further

supporting common underlying pathways. Interestingly, in restricted analyses among participants who initially had an incident cardiovascular event, we observed a markedly strong increase in the Ca predictive accuracy after including baseline methylation data for 80 CpGs annotated to nodes common for Ca, CVD and Ca-CVD DMPs.

In this study we evaluated epigenetic signatures for incident Ca and CVD interrogating as many as ~800000 CpG sites in a population-based study accounting for the simultaneous selection of DMPs. Among the top ten Ca DMPs (Table 2), cg13824555 has been previously associated to cancer (28). The *TMEM132C (29)*, *LGMN* (30), *GINS3* (31), *EPHB3* (32), *TBX18* (33), *SERPINH1* (34) and *WNT3A* (35) genes have been reported differentially expressed in various types of cancer including lung, breast, colorectal or neuroblastoma. Interestingly, many of these genes (*WNT3A*, *SERPINH1*, *TBX18*, *EPHB3* and *GINS3*) have also been reported in CVD studies (36–41). For the top ten CVD-DMPs (Table 3), cg18986048 has been previously reported in association with cancer (42) but not with CVD. Genes such as *EXOSC10* (43), *FMNL3* (44), *TRIM26 (45)* or *MYPN* (46) have been previously associated with CVD.

Among the top ten DMPs associated to Ca-CVD (Table 4), cg08067365 has been previously reported and associated to cancer (47). Four of the top ten DMPs for Ca-CVD were in the *LOC101928804* uncharacterized ncRNA gene and in the *KIR2DL1* gene. The *KIR2DL1* gene encodes a transmembrane glycoprotein expressed by natural killer cells and subsets of T cells, which has been widely studied and related to cancer (48) and hypertension (49). Other genes such as *RNF126* (50,51), *ISG15* (52,53) and *RBFOX1* (54,55) have been related to cell proliferation in cancer (50), metabolic plasticity in pancreatic cancer (52) and mesenchymal cell state in breast cancer (54), as well as to hypertension-induced cardiomyocyte hypertrophy (51), cardiomyopathy (53) and heart failure (55).

A total of 14 DMPs were commonly selected for at least two of the Ca, CVD and Ca-CVD analysis (Table 5) with a noticeable variability in methylation proportions at these genomic sites, but not in the median between groups (Figure 1). These results could be related to epigenetic

plasticity, which has been recently proposed as a potential driver of disease (56). In our results, cancer has the greatest variability/plasticity, compared to non-cases, as expected, which could enable the cancer cells to better adapt and persist in the body. Interestingly, cg14391737 was identified as a Ca-, CVD- and Ca-CVD-DMP and has been previously reported in parent-of-origin effects (57), drugs consumption (58) and smoking (59) -related epigenetic studies. The cg14391737 is in intronic position of *PRSS23*, a gene that encodes a member of trypsin family of serine proteases and has been related to cancer (60) and kidney fibrosis (61). Other overlapping Ca and CVD CpGs such as cg22156456, cg16477091, cg01127300, cg19965693, cg09084200 and cg19764066 have been previously reported as differential methylated in relation to age (62,63), obesity (64,65) and smoking (66,67), which are risk factor for cancer and CVD. DMPs in genes such as *PPP1R3C (68)*, *IRAK2* (69), *PCK2* (70), *IFIH1* (71) and *ESM1* (72), associated with both Ca and CVD, have been previously reported to CVD. DMPs in most other genes associated with both cancer and CVD have been associated to cancer (60,73-76).

The relationship between cancer and cardiovascular diseases can occur in both directions. Cardiovascular diseases can predispose to develop cancer. In a mouse model, the presence of failing hearts resulted in significantly increased intestinal tumor load; the severity of left ventricular dysfunction and fibrotic scar strongly correlated with tumor growth (77). In humans, elevated cardiac and inflammation biomarkers in apparently healthy individuals were predictive of new-onset cancer independently of cancer risk factors (77). Cancer can also be considered as a risk factor for cardiovascular diseases, even years after the diagnosis (78,79). Several cancer treatments, including targeted therapies, produce cardiotoxicity. There are also common genetic and non-genetic risk factors for both cancer and cardiovascular disease (79). For instance, cancer patients are known to suffer from stroke and heart failure complications (80,81). Lung cancer survivors have higher risk of CHD and ischemic stroke compared with non-cancer population (82). Our endpoint-specific analyses found 6 and 4 DMPs common for CHD and HF, respectively, with lung cancer (Figure S1), and 3 common DMPs for stroke and lymphatic, lung and esophageal/stomach cancer,

respectively (Figure S1), after accounting for smoking as a potential confounder. In our study, only found 30 participants with cancer had not developped CVD in the first place, possibly due to the elevated CVD burden in American Indian communities. Our data can thus inform on the increased epigenetic susceptibility of CVD patients to develop Ca, but cannot inform on the potential epigenetic susceptibility of Ca patients to develop CVD.

In the protein-protein interaction network results, hub nodes associated to Ca- CVD- and Ca-CVD-DMPs were genes with well-known functions in cancer and CVD. The *WNT3A* gene encodes a secreted signaling protein implicated in oncogenesis and developmental processes and reported in cancers (e.g., lung, colorectal, breast and others (83,84)) and in CVD (85) such as heart failure (86). The *NKX2-5* gene encodes a homeobox-containing transcription factor which has been involved in papillary thyroid carcinomas (87), cardiomyopathy (88) and congenital heart disease (89). The *TBX3* gene encodes a transcriptional repressor factor that is involved in diseases such as breast cancer (90), gastric cancer (91) or conotruncal heart defects (92). The *EGFR* gene was the most connected node in the network. In our study, it was reported only in association with CVD-DMPs, but it is a gene widely associated to both CVD and Ca. EGFR (also known as ERBB or HER1) is a cell surface protein that binds to epidermal growth factor, which is target of several monoclonal antibodies included in targeted cancer therapy (93). EGFR has been associated to the presence of male-specific Y chromosome mosaicisms of cancer susceptibility (94). In CVD, EGFR has been associated with atherosclerosis (95), hypertension (96), Coronary Artery Disease (97) and others. Further, the networks enrichment analyses confirm that central pathways as Wnt signaling (85,98), AMPK signaling (99,100), PI3K-Akt signaling (100,101), Transcription regulation (100) and Hippo signaling (100,102) are involved in the relationship between cancer and CVDs. Therefore, our findings support the strong and complex biological relationship between cancer and CVD.

Interestingly, Drugbank identified that common DMPs for Ca and CVD were targeted by essential elements. The use of zinc as supplement is controversial given recent evidence suggesting

that elevated zinc exposure may be associated to increased oxidative stress and cardiovascular risk (103,104). Other divalent metals, for instance cadmium which has been related to Ca and CVD in the SHS (105–107) and other study populations (108), and also other metals, can interfere with zinc binding sites. Whether these may be potential precision prevention tools for newly diagnosed CVD patients remained to be clarified in future studies including well-designed intervention trials.

Our study has several limitations. Non-fatal cancer data in the SHS might be incomplete, as no linkage between the SHS cancer data and cancer registry data has been conducted to date. While many of the results were found here for the first time and it is necessary to validate them, experimentally or in another study cohort, the fact that results from Ca only and CVD only differential methylation analysis are consistent with the accumulated evidence in the Ca and CVD fields provide robustness to our findings. Importantly, incident Ca, CVD and Ca-CVD endpoints were analyzed using an informative elastic-net method which enabled avoiding severe multiple-comparison correction methods. Other strengths are related to the fact that the Strong Heart Study (SHS) rich, high quality, resources represent a unique opportunity to study the cardiovascular disease jointly with cancer at epidemiological level, including information of numerous potential confounders, a 30-year follow-up for the assessment of incident endpoint including fatal and non-fatal events, and epigenomic profiles including ~850000 sites across the genome.

Conclusions

The identified epigenetic signatures for Ca and CVD support that underlying common biological mechanisms, including novel and well-established ones, are responsible for Ca and CVD progression. Our drug targets analysis, moreover, support a role for essential elements as potential risk factors. Future studies that evaluate the validity of common epigenetic signatures for Ca and CVD screening can contribute to identify the cancer risk of newly diagnosed CVD patients, enabling precision prevention and control of Ca.

Reference

1. Yeh ETH. Onco-cardiology: the time has come. Tex Heart Inst J. 2011;38(3):246–7.
2. Friedrich CA, Nettleton JA. Cancer : The Atherosclerosis Risk in Communities Study. Circulation. 2014;127(12):1270–5.
3. Koene RJ, Prizment AE, Blaes A, Konety SH. Shared risk factors in cardiovascular disease and cancer. Circulation. 2016;133(11):1104–14.
4. Giza DE, Iliescu G, Hassan S, Marmagkiolis K, Iliescu C. Cancer as a Risk Factor for Cardiovascular Disease. Curr Oncol Rep. 2017;19(6):1–8.
5. Al-Kindi SG, Oliveira GH. Onco-Cardiology: A Tale of Interplay Between 2 Families of Diseases. Mayo Clin Proc. 2016;91(12):1675–7.
6. Hong RA, Iimura T, Sumida KN, Eager RM. Cardio-oncology/onco-cardiology. Clin Cardiol. 2010;33(12):733–7.
7. Aleman BMP, Moser EC, Nuver J, Suter TM, Maraldo M V., Specht L, et al. Cardiovascular disease after cancer therapy. Eur J Cancer Suppl [Internet]. 2014;12(1):18–28. Available from: http://dx.doi.org/10.1016/j.ejcsup.2014.03.002%5Cnhttp://linkinghub.elsevier.com/retrieve/pii/S1359634914000044
8. Michel L, Schadendorf D, Rassaf T. Oncocardiology: new challenges, new opportunities. Herz. 2020 Jun 8;2(8):1–6.
9. Tapia-Vieyra JV, Delgado-Coello B, Mas-Oliva J. Atherosclerosis and Cancer; A Resemblance with Far-reaching Implications. Arch Med Res. 2017;48(1):12–26.
10. Allis CD, Jenuwein T. The molecular hallmarks of epigenetic control. Nat Rev Genet [Internet]. 2016;17(8):487–500. Available from: http://dx.doi.org/10.1038/nrg.2016.59
11. Zoghbi HY, Beaudet AL. Epigenetics and human disease. Cold Spring Harb Perspect Biol. 2016;8(2):1–28.
12. Kinnaird A, Zhao S, Wellen KE, Michelakis ED. Metabolic control of epigenetics in cancer. Nat Rev Cancer [Internet]. 2016;16(11):694–707. Available from: http://dx.doi.org/10.1038/nrc.2016.82
13. van der Harst P, de Windt LJ, Chambers JC. Translational Perspective on Epigenetics in Cardiovascular Disease. J Am Coll Cardiol. 2017;70(5):590–606.
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. 1990;132(6):1141–55.
15. Fortin JP, Triche TJ, Hansen KD. Preprocessing, normalization and integration of the Illumina HumanMethylationEPIC array with minfi. Bioinformatics. 2017;33(4):558–60.

16. Aryee MJ, Jaffe AE, Corrada-Bravo H, Ladd-Acosta C, Feinberg AP, Hansen KD, et al. Minfi: A flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. Bioinformatics. 2014;30(10):1363–9.

17. Triche TJ, Weisenberger DJ, Van Den Berg D, Laird PW, Siegmund KD. Low-level processing of Illumina Infinium DNA Methylation BeadArrays. Nucleic Acids Res. 2013;41(7):1–11.

18. 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. 2012;28(6):882–3.

19. SHS. Center for American Indian Health Research. Strong Heart Study Operations Manual. Phase IV. Volume II: Morbidity and Mortality Surveillance Procedures. University of Oklahoma Health Sciences Center; Oklahoma City, OK: 2001:http://strongheart.ouhsc.ed.

20. Simon N, Friedman J, Hastie T, Tibshirani R. Regularization paths for Cox's proportional hazards model via coordinate descent. J Stat Softw. 2011;39(5):1–13.

21. Friedman J, Hastie T, Tibshirani R. Regularization Paths for Generalized Linear Models via Coordinate Descent. J Stat Softw. 2010;33(1):1–22.

22. Benton MC, Sutherland HG, Macartney-Coxson D, Haupt LM, Lea RA, Griffiths LR. Methylome-wide association study of whole blood DNA in the Norfolk Island isolate identifies robust loci associated with age. Aging (Albany NY). 2017;9(3):753–68.

23. Phipson B, Maksimovic J, Oshlack A. MissMethyl: An R package for analyzing data from Illumina's HumanMethylation450 platform. Bioinformatics. 2016;32(2):286–8.

24. Davis AP, Grondin CJ, Johnson RJ, Sciaky D, McMorran R, Wiegers J, et al. The Comparative Toxicogenomics Database: Update 2019. Nucleic Acids Res. 2019;47(D1):D948–54.

25. 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;47(D1):D607–13.

26. 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 1;13(11):2498–504.

27. 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. 2011;39(SUPPL. 1):1035–41.

28. Luo Y, Wong CJ, Kaz AM, Dzieciatkowski S, Carter KT, Morris SM, et al. Differences in DNA methylation signatures reveal multiple pathways of progression from adenoma to colorectal cancer. Gastroenterology. 2014;147(2).

29. De Almeida BP, Apolónio JD, Binnie A, Castelo-Branco P. Roadmap of DNA methylation in breast cancer identifies novel prognostic biomarkers. BMC Cancer. 2019;19(1):1–12.

30. Zhen Y, Chunlei G, Wenzhi S, Shuangtao Z, Na L, Rongrong W, et al. Clinicopathologic significance of legumain overexpression in cancer: A systematic review and meta-analysis. Sci Rep [Internet]. 2015;5(October):1–9. Available from: http://dx.doi.org/10.1038/srep16599

31. Tauchi S, Sakai Y, Fujimoto S, Ogawa H, Tane S, Hokka D, et al. Psf3 is a prognostic biomarker in lung adenocarcinoma: A larger trial using tissue microarrays of 864 consecutive resections. Eur J Cardio-thoracic Surg. 2016;50(4):758–64.

32. Jang BG, Kim HS, Bae JM, Kim WH, Hyun CL, Kang GH. Expression profile and prognostic significance of EPHB3 in colorectal cancer. Biomolecules. 2020;10(4):1–18.

33. Su Y, Subedee A, Bloushtain-Qimron N, Savova V, Krzystanek M, Li L, et al. Somatic Cell Fusions Reveal Extensive Heterogeneity in Basal-like Breast Cancer. Cell Rep [Internet]. 2015;11(10):1549–63. Available from: http://dx.doi.org/10.1016/j.celrep.2015.05.011

34. Zhang X, Kang X, Jin L, Bai J, Zhang H, Liu W, et al. ABCC9, NKAPL, and TMEM132C are potential diagnostic and prognostic markers in triple-negative breast cancer. Cell Biol Int. 2020 Jul;cbin.11406.

35. Pashirzad M, Fiuji H, Khazei M, Moradi-Binabaj M, Ryzhikov M, Shabani M, et al. Role of Wnt3a in the pathogenesis of cancer, current status and prospective. Mol Biol Rep. 2019;46(5):5609–16.

36. Zhang Y, Zhang L, Fan X, Yang W, Yu B, Kou J, et al. Captopril attenuates TAC-induced heart failure via inhibiting Wnt3a/β-catenin and Jak2/Stat3 pathways. Biomed Pharmacother [Internet]. 2019;113(February):108780. Available from: https://doi.org/10.1016/j.biopha.2019.108780

37. Li X, Ren Y, Sorokin V, Poh KK, Ho HH, Lee CN, et al. Quantitative profiling of the rat heart myoblast secretome reveals differential responses to hypoxia and re-oxygenation stress. J Proteomics [Internet]. 2014;98(Mi):138–49. Available from: http://dx.doi.org/10.1016/j.jprot.2013.12.025

38. Lacey M, Baribault C, Ehrlich KC, Ehrlich M. Atherosclerosis-associated differentially methylated regions can reflect the disease phenotype and are often at enhancers. Atherosclerosis. 2019;280(November 2018):183–91.

39. Braitsch CM, Kanisicak O, van Berlo JH, Molkentin JD, Yutzey KE. Differential expression of embryonic epicardial progenitor markers and localization of cardiac fibrosis in adult ischemic injury and hypertensive heart disease. J Mol Cell Cardiol [Internet]. 2013;65:108–19. Available from: http://dx.doi.org/10.1016/j.yjmcc.2013.10.005

40. Adams RH, Wilkinson GA, Weiss C, Diella F, Gale NW, Deutsch U, et al. Roles of ephrinB ligands and EphB receptors in cardiovascular development: Demarcation of arterial/venous domains, vascular morphogenesis, and sprouting angiogenesis. Genes Dev. 1999;13(3):295–306.

41. Milan DJ, Kim AM, Winterfield JR, Jones IL, Pfeufer A, Sanna S, et al. Drug-sensitized zebrafish screen identifies multiple genes, including GINS3, as regulators of myocardial

repolarization. Circulation. 2009;120(7):553–9.

42. Giri AK, Aittokallio T. DNMT inhibitors increase methylation at subset of CpGs in colon, bladder, lymphoma, breast, and ovarian, cancer genome. bioRxiv. 2018;

43. Abdulrahim JW, Kwee LC, Grass E, Siegler IC, Williams R, Karra R, et al. Epigenome-Wide Association Study for All-Cause Mortality in a Cardiovascular Cohort Identifies Differential Methylation in Castor Zinc Finger 1 (CASZ1). J Am Heart Assoc. 2019;8(21).

44. Rosado M, Barber CF, Berciu C, Feldman S, Birren SJ, Nicastro D, et al. Critical roles for multiple formins during cardiac myofibril development and repair. Mol Biol Cell. 2014;25(6):811–27.

45. Li X, Wang X, Liu YS, Wang XD, Zhou J, Zhou H. Downregulation of miR-3568 Protects Against Ischemia/Reperfusion-Induced Cardiac Dysfunction in Rats and Apoptosis in H9C2 Cardiomyocytes Through Targeting TRIM62. Front Pharmacol. 2020;11(February):1–11.

46. Purevjav E, Arimura T, Augustin S, Huby AC, Takagi K, Nunoda S, et al. Molecular basis for clinical heterogeneity in inherited cardiomyopathies due to myopalladin mutations. Hum Mol Genet. 2012;21(9):2039–53.

47. Sartor MA, Dolinoy DC, Jones TR, Colacino JA, Prince MEP, Carey TE, et al. Genome-wide methylation and expression differences in HPV(+) and HPV(-) squamous cell carcinoma cell lines are consistent with divergent mechanisms of carcinogenesis. Epigenetics. 2011;6(6):777–87.

48. Guillamón CF, Gimeno L, Server G, Martínez-Sánchez M V., Escudero JF, López-Cubillana P, et al. Immunological Risk Stratification of Bladder Cancer Based on Peripheral Blood Natural Killer Cell Biomarkers. Eur Urol Oncol. 2019;1–10.

49. Ormiston ML, Chang C, Long LL, Soon E, Jones D, Machado R, et al. Impaired natural killer cell phenotype and function in idiopathic and heritable pulmonary arterial hypertension. Circulation. 2012;126(9):1099–109.

50. Zhi X, Zhao D, Wang Z, Zhou Z, Wang C, Chen W, et al. E3 ubiquitin ligase RNF126 promotes cancer cell proliferation by targeting the tumor suppressor p21 for ubiquitin-mediated degradation. Cancer Res. 2013;73(1):385–94.

51. Huang CY, Lee FL, Peng SF, Lin KH, Chen RJ, Ho TJ, et al. HSF1 phosphorylation by ERK/GSK3 suppresses RNF126 to sustain IGF-IIR expression for hypertension-induced cardiomyocyte hypertrophy. J Cell Physiol. 2018;233(2):979–89.

52. Alcalá S, Sancho P, Martinelli P, Navarro D, Pedrero C, Martín-Hijano L, et al. ISG15 and ISGylation is required for pancreatic cancer stem cell mitophagy and metabolic plasticity. Nat Commun [Internet]. 2020;11(1):1–17. Available from: http://dx.doi.org/10.1038/s41467-020-16395-2

53. Rahnefeld A, Klingel K, Schuermann A, Diny NL, Althof N, Lindner A, et al. Ubiquitin-like protein ISG15 (Interferon-Stimulated Gene of 15 kDa) in host defense against heart failure in a mouse model of virus-induced cardiomyopathy. Circulation. 2014;130(18):1589–600.

54. Li J, Choi PS, Chaffer CL, Labella K, Hwang JH, Giacomelli AO, et al. An alternative splicing switch in FLNB promotes the mesenchymal cell state in human breast cancer. Elife. 2018;7:1–28.

55. Gao C, Ren S, Lee JH, Qiu J, Chapski DJ, Rau CD, et al. RBFox1-mediated RNA splicing regulates cardiac hypertrophy and heart failure. J Clin Invest. 2016;126(1):195–206.

56. Feinberg AP, Irizarry RA. Stochastic epigenetic variation as a driving force of development, evolutionary adaptation, and disease. Proc Natl Acad Sci U S A. 2010;107(SUPPL. 1):1757–64.

57. Zeng Y, Amador C, Xia C, Marioni R, Sproul D, Walker RM, et al. Parent of origin genetic effects on methylation in humans are common and influence complex trait variation. Nat Commun. 2019;10(1):1–13.

58. Osborne AJ, Pearson JF, Noble AJ, Gemmell NJ, Horwood LJ, Boden JM, et al. Genome-wide DNA methylation analysis of heavy cannabis exposure in a New Zealand longitudinal cohort. Transl Psychiatry. 2020;10(1).

59. Domingo-Relloso 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. 2020;128(6).

60. Han B, Yang Y, Chen J, He X, Lv N, Yan R. PRSS23 knockdown inhibits gastric tumorigenesis through EIF2 signaling. Pharmacol Res. 2019;142(February):50–7.

61. LeBleu VS, Teng Y, O'Connell JT, Charytan D, Müller GA, Müller CA, et al. Identification of human epididymis protein-4 as a fibroblast-derived mediator of fibrosis. Nat Med. 2013 Feb 27;19(2):227–31.

62. Florath I, Butterbach K, Müller H, Bewerunge-hudler M, Brenner H. Cross-sectional and longitudinal changes in DNA methylation with age: An epigenome-wide analysis revealing over 60 novel age-associated CpG sites. Hum Mol Genet. 2014;23(5):1186–201.

63. Naue J, Hoefsloot HCJ, Mook ORF, Rijlaarsdam-Hoekstra L, van der Zwalm MCH, Henneman P, et al. Chronological age prediction based on DNA methylation: Massive parallel sequencing and random forest regression. Forensic Sci Int Genet [Internet]. 2017;31:19–28. Available from: http://dx.doi.org/10.1016/j.fsigen.2017.07.015

64. Benton MC, Johnstone A, Eccles D, Harmon B, Hayes MT, Lea RA, et al. An analysis of DNA methylation in human adipose tissue reveals differential modification of obesity genes before and after gastric bypass and weight loss. Genome Biol. 2015;16(1):1–21.

65. Xu J, Bao X, Peng Z, Wang L, Du L, Niu W, et al. Comprehensive analysis of genome-wide DNA methylation across human polycystic ovary syndrome ovary granulosa cell. Oncotarget. 2016;7(19):27899–909.

66. Breitling LP, Yang R, Korn B, Burwinkel B, Brenner H. Tobacco-smoking-related differential DNA methylation: 27K discovery and replication. Am J Hum Genet. 2011;88(4):450–7.

67. Gao X, Zhang Y, Saum KU, Schöttker B, Breitling LP, Brenner H. Tobacco smoking and smoking-related DNA methylation are associated with the development of frailty among older adults. Epigenetics. 2017;12(2):149–56.

68. Wu P, Zhang Z, Ma G, Li J, Zhou W. Transcriptomics and metabolomics reveal the cardioprotective effect of Compound Danshen tablet on isoproterenol-induced myocardial injury in high-fat-diet fed mice. J Ethnopharmacol. 2020;246(August 2019):112210.

69. Su Q, Lv X, Ye Z, Sun Y, Kong B, Qin Z, et al. The mechanism of miR-142-3p in coronary microembolization-induced myocardiac injury via regulating target gene IRAK-1. Cell Death Dis. 2019;10(2).

70. Ma H, Yu S, Liu X, Zhang Y, Fakadej T, Liu Z, et al. Lin28a Regulates Pathological Cardiac Hypertrophic Growth Through Pck2-Mediated Enhancement of Anabolic Synthesis. Circulation. 2019;139(14):1725–40.

71. Buskiewicz IA, Huber SA. The almighty for myocardium. Circ Hear Fail. 2013;6(2):153–5.

72. Sun H, Fang F, Li K, Zhang H, Zhang M, Zhang L, et al. Circulating ESM-1 levels are correlated with the presence of coronary artery disease in patients with obstructive sleep apnea. Respir Res. 2019;20(1):1–9.

73. Liu J, Chen Y, Huang Q, Liu W, Ji X, Hu F, et al. IRAK2 counterbalances oncogenic Smurf1 in colon cancer cells by dictating ER stress. Cell Signal. 2018 Aug;48:69–80.

74. Luo S, Li Y, Ma R, Liu J, Xu P, Zhang H, et al. Downregulation of PCK2 remodels tricarboxylic acid cycle in tumor-repopulating cells of melanoma. Oncogene. 2017 Jun 6;36(25):3609–17.

75. van Agthoven T, Sieuwerts AM, Meijer D, Meijer-van Gelder ME, van Agthoven TLA, Sarwari R, et al. Selective recruitment of breast cancer anti-estrogen resistance genes and relevance for breast cancer progression and tamoxifen therapy response. Endocr Relat Cancer. 2010 Mar;17(1):215–30.

76. Chen M-B, Liu Y-Y, Cheng L-B, Lu J-W, Zeng P, Lu P-H. AMPKα phosphatase Ppm1E upregulation in human gastric cancer is required for cell proliferation. Oncotarget. 2017 May 9;8(19):31288–96.

77. Meijers WC, Maglione M, Bakker SJL, Oberhuber R, Kieneker LM, de Jong S, et al. Heart Failure Stimulates Tumor Growth by Circulating Factors. Circulation. 2018 Aug 14;138(7):678–91.

78. Sturgeon KM, Deng L, Bluethmann SM, Zhou S, Trifiletti DM, Jiang C, et al. A population-based study of cardiovascular disease mortality risk in US cancer patients. Eur Heart J. 2019;40(48):3889–97.

79. Campia U, Moslehi JJ, Amiri-Kordestani L, Barac A, Beckman JA, Chism DD, et al. Cardio-Oncology: Vascular and Metabolic Perspectives: A Scientific Statement From the American Heart Association. Circulation. 2019;139(13):E579–602.

80. Zaorsky NG, Zhang Y, Tchelebi LT, Mackley HB, Chinchilli VM, Zacharia BE. Stroke among cancer patients. Nat Commun [Internet]. 2019;10(1). Available from: http://dx.doi.org/10.1038/s41467-019-13120-6

81. Samejima Y, Iuchi A, Kanai T, Noda Y, Nasu S, Tanaka A, et al. Development of severe heart failure in a patient with squamous non-small-cell lung cancer during nivolumab treatment. Intern Med. 2020;

82. Yoon DW, Shin DW, Cho JH, Yang JH, Jeong SM, Han K, et al. Increased risk of coronary heart disease and stroke in lung cancer survivors: A Korean nationwide study of 20,458 patients. Lung Cancer [Internet]. 2019;136(May):115–21. Available from: https://doi.org/10.1016/j.lungcan.2019.08.025

83. Pashirzad M, Fiuji H, Khazei M, Moradi-Binabaj M, Ryzhikov M, Shabani M, et al. Role of Wnt3a in the pathogenesis of cancer, current status and prospective. Mol Biol Rep. 2019 Oct 24;46(5):5609–16.

84. Liu G, Wang P, Zhang H. MiR-6838-5p suppresses cell metastasis and the EMT process in triple-negative breast cancer by targeting WNT3A to inhibit the Wnt pathway. J Gene Med. 2019 Dec 3;21(12).

85. Foulquier S, Daskalopoulos EP, Lluri G, Hermans KCM, Deb A, Blankesteijn WM. WNT signaling in cardiac and vascular disease. Pharmacol Rev. 2018;70(1):68–141.

86. Zhang Y, Zhang L, Fan X, Yang W, Yu B, Kou J, et al. Captopril attenuates TAC-induced heart failure via inhibiting Wnt3a/β-catenin and Jak2/Stat3 pathways. Biomed Pharmacother. 2019 May;113:108780.

87. Cortez R, Penha C, Buexm LA, Rodrigues FR, Castro TP De, Santos MCS, et al. NKX2 . 5 is expressed in papillary thyroid carcinomas and regulates differentiation in thyroid cells. BMC Cancer. 2018;18(498):1–12.

88. Xu J-H, Gu J-Y, Guo Y-H, Zhang H, Qiu X-B, Li R-G, et al. Prevalence and Spectrum of NKX2-5 Mutations Associated With Sporadic Adult-Onset Dilated Cardiomyopathy. Int Heart J. 2017;58(4):521–9.

89. Khatami M, Mazidi M, Taher S, Heidari M, Hadadzadeh M. Novel Point Mutations in the NKX2.5 Gene in Pediatric Patients with Non-Familial Congenital Heart Disease. Medicina (B Aires). 2018 Jun 19;54(3):46.

90. Krstic M, Kolendowski B, Cecchini MJ, Postenka CO, Hassan HM, Andrews J, et al. TBX3 promotes progression of pre-invasive breast cancer cells by inducing EMT and directly up-regulating SLUG. J Pathol. 2019 Jun 8;248(2):191–203.

91. Miao Z-F, Liu X-Y, Xu H-M, Wang Z-N, Zhao T-T, Song Y-X, et al. Tbx3 overexpression in human gastric cancer is correlated with advanced tumor stage and nodal status and promotes cancer cell growth and invasion. Virchows Arch. 2016 Nov 24;469(5):505–13.

92. Xie H, Zhang E, Hong N, Fu Q, Li F, Chen S, et al. Identification of TBX2 and TBX3 variants in patients with conotruncal heart defects by target sequencing. Hum Genomics. 2018 Dec 17;12(1):44.

93. Wieduwilt MJ, Moasser MM. The epidermal growth factor receptor family: Biology driving targeted therapeutics. Cell Mol Life Sci. 2008 May 9;65(10):1566–84.

94. Cáceres A, Jene A, Esko T, Pérez-Jurado LA, González JR. Extreme Downregulation of Chromosome Y and Cancer Risk in Men. JNCI J Natl Cancer Inst. 2020 Jan 16;7:djz232.

95. Zeboudj L, Maître M, Guyonnet L, Laurans L, Joffre J, Lemarie J, et al. Selective EGF-Receptor Inhibition in CD4+ T Cells Induces Anergy and Limits Atherosclerosis. J Am Coll Cardiol. 2018 Jan;71(2):160–72.

96. Dahal BK, Cornitescu T, Tretyn A, Pullamsetti SS, Kosanovic D, Dumitrascu R, et al. Role of Epidermal Growth Factor Inhibition in Experimental Pulmonary Hypertension. Am J Respir Crit Care Med. 2010 Jan 15;181(2):158–67.

97. Wang S, He W, Wang C. MiR-23a Regulates the Vasculogenesis of Coronary Artery Disease by Targeting Epidermal Growth Factor Receptor. Cardiovasc Ther. 2016 Aug;34(4):199–208.

98. Zhan T, Rindtorff N, Boutros M. Wnt signaling in cancer. Oncogene [Internet]. 2017;36(11):1461–73. Available from: http://dx.doi.org/10.1038/onc.2016.304

99. Carling D. AMPK signalling in health and disease. Curr Opin Cell Biol [Internet]. 2017;45:31–7. Available from: http://dx.doi.org/10.1016/j.ceb.2017.01.005

100. Sever R, Brugge JS. Signal Transduction in Cancer. Cold Spring Harb Perspect Med. 2015;5(4):a006098.

101. Aoyagi T, Matsui T. Phosphoinositide-3 Kinase Signaling in Cardiac Hypertrophy and Heart Failure. Curr Pharm Des. 2011;17(18):1818–24.

102. Zhou Q, Li L, Zhao B, Guan K. The Hippo Pathway in Heart Development, Regeneration, and Diseases. Circ Res [Internet]. 2015 Apr 10;116(8):1431–47. Available from: https://www.ahajournals.org/doi/10.1161/CIRCRESAHA.116.303311

103. 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. 2019 Feb;123:171–80.

104. Domingo-Relloso A, Grau-Perez M, Briongos-Figuero L, Gomez-Ariza JL, Garcia-Barrera T, Dueñas-Laita A, et al. The association of urine metals and metal mixtures with cardiovascular incidence in an adult population from Spain: the Hortega Follow-Up Study. Int J Epidemiol. 2019 Dec 1;48(6):1839–49.

105. 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. 2014 Apr;122(4):363–70.

106. Tellez-Plaza M, Guallar E, Fabsitz RR, Howard B V, Umans JG, Francesconi KA, et al. Cadmium exposure and incident peripheral arterial disease. Circ Cardiovasc Qual Outcomes. 2013 Nov;6(6):626–33.

107. Tellez-Plaza M, Guallar E, Howard B V., Umans JG, Francesconi KA, Goessler W, et al. Cadmium Exposure and Incident Cardiovascular Disease. Epidemiology. 2013 May;24(3):421–9.

108. Nigra AE, Ruiz-Hernandez A, Redon J, Navas-Acien A, Tellez-Plaza M. Environmental Metals and Cardiovascular Disease in Adults: A Systematic Review Beyond Lead and Cadmium. Curr Environ Heal Reports. 2016 Dec 25;3(4):416–33.

TABLES

Table 1. Cohorts characterization. For continuous variables the median interquartile range are shown and for categorical variables percentages are shown.

Variables	Non cases	Incident Cancer (Ca) no CVD cases	Incident CVD no Cancer cases	Incident Ca-CVD cases
Number of patients	944	277	823	142
Age	52.9 (48.0, 59.9)	56.2 (50.4, 63.8)	56.6 (50.6, 63.7)	56.9 (50.9, 63.5)
Female %	61.1	54.2	56.5	49.3
Arizona %	17.1	10.0	9.7	3.5
Oklahoma %	47.7	34.8	38.6	33.1
North/South Dakota %	35.2	55.1	51.7	63.4
Never smoking %	33.0	22.7	27.6	21.1
Former smoking %	32.3	31.0	32.4	31.7
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)
Micro albuminuria %	11.6	16.9	19.5	18.3
Macro albuminuria %	5.1	3.8	10.4	5.6

Table 2. Top ten Ca-DMPs selected by the elastic-net model. The CpG position column corresponds to the GRCh37/hg19 human genome version.

CpG	CpG position	HR (95% CI)	P-value	Location	Gene	Description
cg13824555	Chr12:128753138	2.05 (1.61-2.61)	6.89E-09	Intron	TMEM132C	Transmembrane protein
cg23327070	Chr14:93214896	1.31 (1.19-1.45)	5.92E-08	5’ UTR	LGMN	Hydrolysis of asparaginyl bonds
cg12863924	Chr16:58443482	0.66 (0.56-0.77)	1.78E-07	Intergenic	GINS3*	Initiation of DNA replication and replisome progression
cg00229368	Chr3:184279621	1.50 (1.28-1.75)	3.43E-07	5’ UTR	EPHB3	Receptor for ephrin-B family members
cg01839396	Chr6:85484450	2.01 (1.52-2.64)	7.50E-07	Intergenic	TBX18*	T-box transcription factors
cg27596068	Chr11:75272301	0.55 (0.44-0.70)	7.75E-07	Promoter (TSS1500)	SERPINH1	Serpin superfamily of serine proteinase inhibitors
cg19764066	Chr2:3819033	0.51 (0.39-0.68)	2.38E-06	Intergenic	ALLC*	Uric acid degradation
cg05081800	Chr1:228227828	0.50 (0.37-0.67)	3.54E-06	Intron	WNT3A	Oncogenesis and developmental processes
cg03734285	Chr12:65222734	0.64 (0.53-0.77)	3.60E-06	Intergenic	TBC1D30*	GTPase-activating protein
cg06761719	Chr6:27280190	1.89 (1.44-2.48)	3.81E-06	Promoter (TSS200)	POM121L2	Transmembrane nucleoporin

*Intergenic CpGs assigned to the closest gene; HR: Hazard Ratio.

Models adjusted for age, sex, BMI, smoking status, study center, cell counts and five genetic PCs.

Table 3. Top ten CVD-DMPs selected by the elastic-net model. The position column corresponds to the GRCh37/hg19 human genome version.

CpG	CpG position	HR (95% CI)	P-value	Location	Gene	Description
cg21025681	Chr1:11134131	1.79 (1.42-2.25)	6.77E-07	Intron	EXOSC10	Exosome component
cg14734341	Chr11:133678993	1.62 (1.33-1.97)	1.82E-06	Intron	LOC646522	Long intergenic non-protein coding RNA
cg15843572	Chr12:133249005	1.62 (1.33- 1.97)	1.85E-06	Intron	POLE	Catalytic subunit of DNA polymerase epsilon
cg02191320	Chr2:55186282	1.71 (1.37-2.14)	2.57E-06	Exon	EML6	Unknown
cg09797037	Chr1:11132717	1.59 (1.31-1.93)	3.55E-06	Intron	EXOSC10	Exosome component
cg18728406	Chr12:50066533	1.53 (1.28-1.84)	4.12E-06	Intron	FMNL3	Formin homology 2 domain
cg00621366	Chr1:226322846	1.52 (1.27-1.82)	5.79E-06	Intergenic	ACBD3*	Maintenance of Golgi structure and function
cg24701662	Chr6:30166168	1.53 (1.27-1.85)	1.07E-05	Intron	TRIM26	DNA-binding activity
cg18986048	Chr10:69913749	1.49 (1.25-1.78)	1.11E-05	Intron	MYPN	Interacts with nebulette in cardiac muscle and alpha-actinin
cg15026376	Chr1:161238112	1.57 (1.28-1.92)	1.19E-05	Intron	PCP4L1	Unknown, purkinje cell protein

*Intergenic CpGs assigned to the closest gene; HR: Hazard Ratio.

Models adjusted for age, sex, BMI, smoking status, study center, cell counts, five genetic PCs, LDL cholesterol, HDL cholesterol, diabetes, hypertension, systolic blood pressure and albuminuria.

Table 4. Top ten Ca-CVD-DMPs selected by the elastic-net model. The position column corresponds to the GRCh37/hg19 human genome version.

CpG	CpG position	HR (95% CI)	P-value	Location	Gene	Description
cg25352098	Chr19:55281126	0.76 (0.70-0.84)	2.84E-09	Exon	KIR2DL1; LOC101928804	Killer cell immunoglobulin-like receptors; uncharacterized ncRNA
cg04666037	Chr19:665899	0.62 (0.52-0.73)	2.15E-08	Intergenic	RNF126*	RING finger domain
cg08067365	Chr1:948893	1.44 (1.26-1.65)	6.18E-08	Exon	ISG15	Ubiquitin-like protein involved in cell-to-cell signaling and others
cg01228290	Chr11:91130448	0.21 (0.12-0.38)	7.76E-08	Intergenic	MIR4490*	Post-transcriptional regulation of gene expression
cg24821564	Chr19:55280672	0.70 (0.62-0.80)	1.43E-07	Promoter (TSS1500)	KIR2DL1	Killer cell immunoglobulin-like receptors
cg01387220	Chr19:55281274	0.84 (0.79-0.90)	1.58E-07	Exon	KIR2DL1; LOC101928804	Killer cell immunoglobulin-like receptors; uncharacterized ncRNA
cg03391464	Chr16:7374542	0.27 (0.17-0.45)	2.65E-07	Intron	A2BP1 (RBFOX1)	RNA-binding protein
cg16018314	Chr6:123100582	0.31 (0.19-0.48)	3.92E-07	Intron	FABP7	Bind long-chain fatty acids
cg04309942	Chr3:81267495	0.29 (0.18-0.47)	5.13E-07	Intergenic	GBE1*	Glycogen branching enzyme
cg13728299	Chr19:55281133	0.74 (0.66-0.83)	7.32E-07	Promoter (TSS200); Exon	KIR2DL1; LOC101928804	Killer cell immunoglobulin-like receptors; uncharacterized ncRNA

*Intergetic CpGs assigned to the closest gene; HR: Hazard Ratio.

Models adjusted for age, sex, BMI, smoking status, study center, cell counts, five genetic PCs, LDL cholesterol, HDL cholesterol, diabetes, hypertension, systolic blood pressure and albuminuria.

Table 5. DMPs in common for cancer and CVD. The common DMPs in the overlap between cancer, CVD and Ca-CVD, showed in the Venn diagram of the figure 1. The position column corresponds to the GRCh37/hg19 human genome version. All DMPs were significant with a nominal p-value <0.05. The "-" symbol indicates that the CpG was not reported in the group of patients.

CpG	CpG position	Location	Gene	CVD HR (95% CI)	Cancer HR (95% CI)	Ca-CVD HR (95% CI)
cg14391737	Chr11:86513429	Intron	PRSS23	0.66 (0.53-0.83)	0.58 (0.40-0.86)	0.36 (0.20-0.63)
cg22156456	Chr17:39844239	Promoter (TSS1500)$^{+}$	EIF1	-	0.63 (0.46-0.86)	0.49 (0.32-0.77)
cg00688297	Chr8:145752292	5'UTR^{-}; 3'UTR^{-}	LRRC24; MGC70857	1.31 (1.10-1.55)	-	1.66 (1.12-2.46)
cg02935487	Chr10:93432326	Intergenic	PPP1R3C*	0.77 (0.64-0.93)	-	0.49 (0.32-0.76)
cg10106965	Chr3:10206258	Promoter (TSS1500)	IRAK2	0.80 (0.67-0.96)	-	0.39 (0.25-0.60)
cg01127300	Chr22:38614796	Intergenic	TMEM184B*	0.75 (0.62-0.90)	-	0.50 (0.33-0.78)
cg16138677	Chr14:24564438	Intron	PCK2; NRL	0.82 (0.70-0.96)	-	0.49 (0.31-0.77)
cg19965693	Chr2:163175743	Promoter (TSS1500)	IFIH1	0.70 (0.57-0.85)	-	0.40 (0.24-0.68)
cg25535569	Chr5:54163217	Intergenic	ESM1*	0.78 (0.64-0.96)	-	0.37 (0.22-0.64)
cg00549219	Chr5:177370476	Intergenic	LOC728554*	0.74 (0.61-0.91)	-	0.42 (0.26-0.67)
cg07503069	Chr2:241084049	Intergenic	OTOS*	0.80 (0.66-0.98)	0.60 (0.42-0.83)	-

cg09084200	Chr11:134095863	Promoter (TSS1500); Intron^{+}	NCAPD3; VPS26B	0.79 (0.66-0.95)	0.61 (0.44-0.85)	-
cg16477091	Chr17:56833000	Promoter (TSS1500)	PPM1E	1.32 (1.10-1.59)	1.78 (1.29-2.45)	-
cg19764066	Chr2:3819033	Intergenic	ALLC*	0.77 (0.64-0.92)	0.51 (0.39-0.68)	-

*Intergetic CpGs assigned to the more closest gene; $^{+}$The gene is in forward strand; $^{-}$The gene is in reverse strand; HR: Hazard Ratio

Table 6. Genes reported in the drug target analysis using XXX.

DrugBank ID	Drug name	Description and/or indication	Target (UniProt ID)
DB01593	Zinc	Zinc imbalance may be associated with neuronal damage associated with traumatic brain injury, stroke, and seizures. Zinc can be used for the treatment and prevention of zinc deficiency/its consequences, including stunted growth and acute diarrhea in children, and slowed wound healing. It is also used for boosting the immune system, treating common cold and recurrent ear infections, and preventing lower respiratory tract infections	JUP (P14923); TP73 (O15350)
DB14487	Zinc acetate	Zinc can be used for the treatment and prevention of zinc deficiency/its consequences, including stunted growth and acute diarrhea in children, and slowed wound healing. It is also utilized for boosting the immune system, treating the common cold and recurrent ear infections, as well as preventing lower respiratory tract infections	JUP (P14923); TP73 (O15350)
DB14533	Zinc chloride	Zinc chloride injections are indicated for use total parenteral nutrition to maintain zinc serum levels and prevent deficiency syndromes	TP73 (O15350)
DB14548	Zinc sulfate	Not Available	TP73 (O15350)
DB04941	Crofelemer	Crofelemer, previously known as investigational drug SP-303, is a proanthocyanidin purified from the bark latex of the Amazonian Croton tree Croton lechleri. It is marketed under the name Fulyzaq and indicated for the symptomatic treatment of non-infectious diarrhea in adult patients with HIV/AIDS who are taking antiretroviral therapy.	ANO1 (Q5XXA6)
DB00121	Biotin	A water-soluble, enzyme co-factor present in minute amounts in every living cell. It occurs mainly bound to proteins or polypeptides and is abundant in liver, kidney, pancreas, yeast, and milk. For nutritional supplementation, also for treating dietary shortage or imbalance	HLCS (P50747)
DB00787	Acyclovir	Acyclovir is a nucleotide analog antiviral used to treat herpes simplex, varicella zoster, herpes zoster, herpes labialis, and acute herpetic keratitis. An acyclovir topical cream is indicated to treat recurrent herpes labialis in immunocompetent patients 12 years and older. Acyclovir oral tablets, capsules, and suspensions are indicated to treat herpes zoster, genital herpes, and chickenpox. An acyclovir topical ointment is	PCK2 (Q16822)

	indicated to treat initial genital herpes and limited non-life-threatening mucocutaneous herpes simplex in immunocompromised patients. Acyclovir cream with hydrocortisone is indicated to treat recurrent herpes labialis, and shortening lesion healing time in patients 6 years and older. Acyclovir buccal tablet is indicated for recurrent herpes labialis. An acyclovir ophthalmic ointment is indicated to treat acute herpetic keratitis.

FIGURES

Figure 1. Workflow for the analysis of the association of blood DNA methylation with incident cancer and cardiovascular disease. The 2186 participants of the Strong Heart Study were grouped as non-cases, incident CVD with no cancer (Ca), incident Ca with no CVD, and incident Ca-CVD. The 3 groups of cases were contrasted with non-cases in a genomic Cox elastic-net assessment of ~790000 CpG sites included in the Infinium MethylationEPIC BeadChip. Of these, 1092 DMPs (nominal p-value < 0.05) were associated with either incident CVD, Ca and Ca-CVD, and 14 DMPs were in common for both cancer and CVD (see violin plot). The DMPs were included in a GO terms enrichment analyses. In addition, the protein-coding genes associated with the CpG sites were included in a protein-protein interaction network.

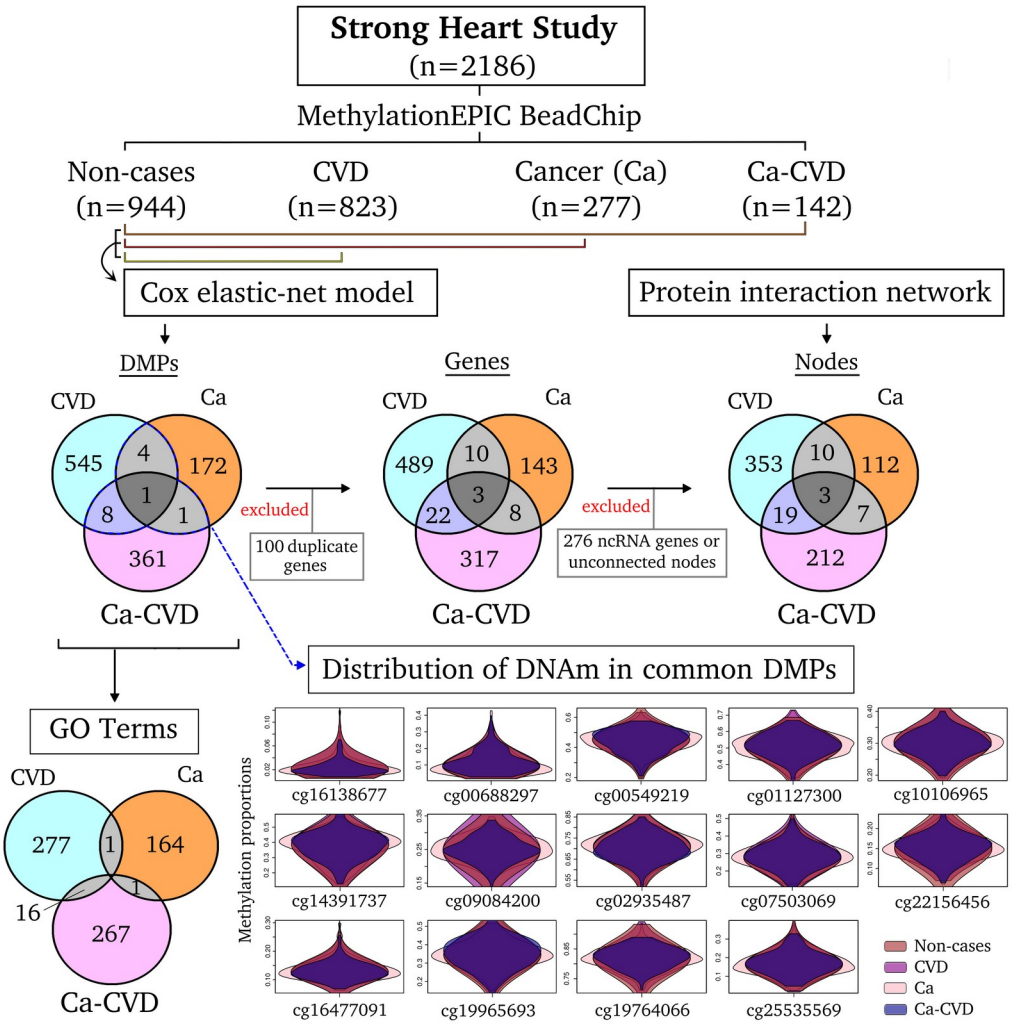
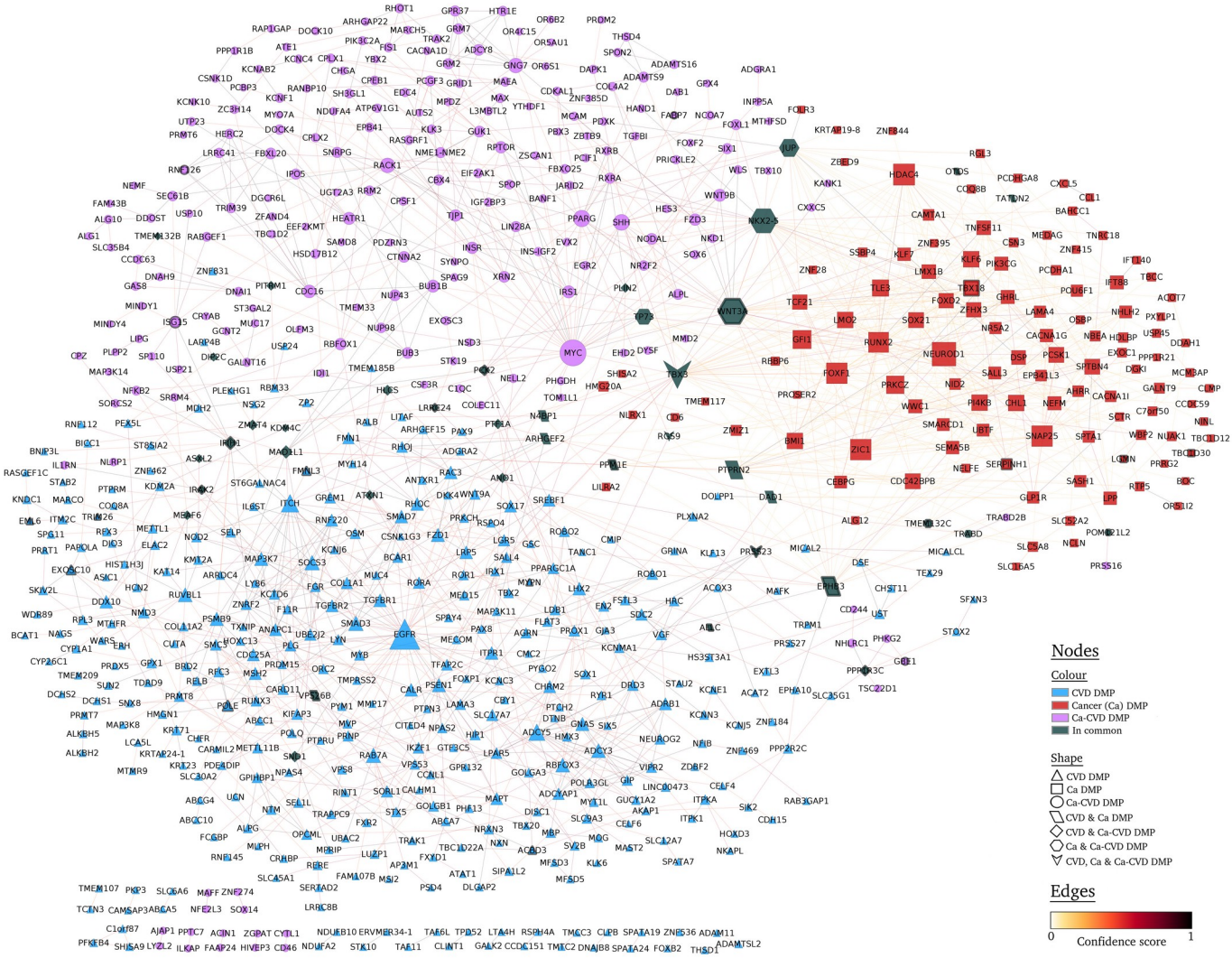


Figure 2. Protein-protein interaction network for cancer and cardiovascular diseases. The network is displayed by Cytoscape using the yFiles organic layout and contains 716 nodes and 1622 interactions. The nodes correspond to protein-coding genes with DMPs associated to cancer and cardiovascular diseases. The size of the nodes is proportional to the number of connections, the color is according the disease and the shape show the overlap between diseases. Increasingly darker solid edge lines indicate protein interaction with increasingly confidence scores. The interaction and its confidence score were assigned by STRING database. Thick-edged nodes indicate the CpGs reported in Tables 2, 3, 4 and 5.



SUPPLEMENTARY

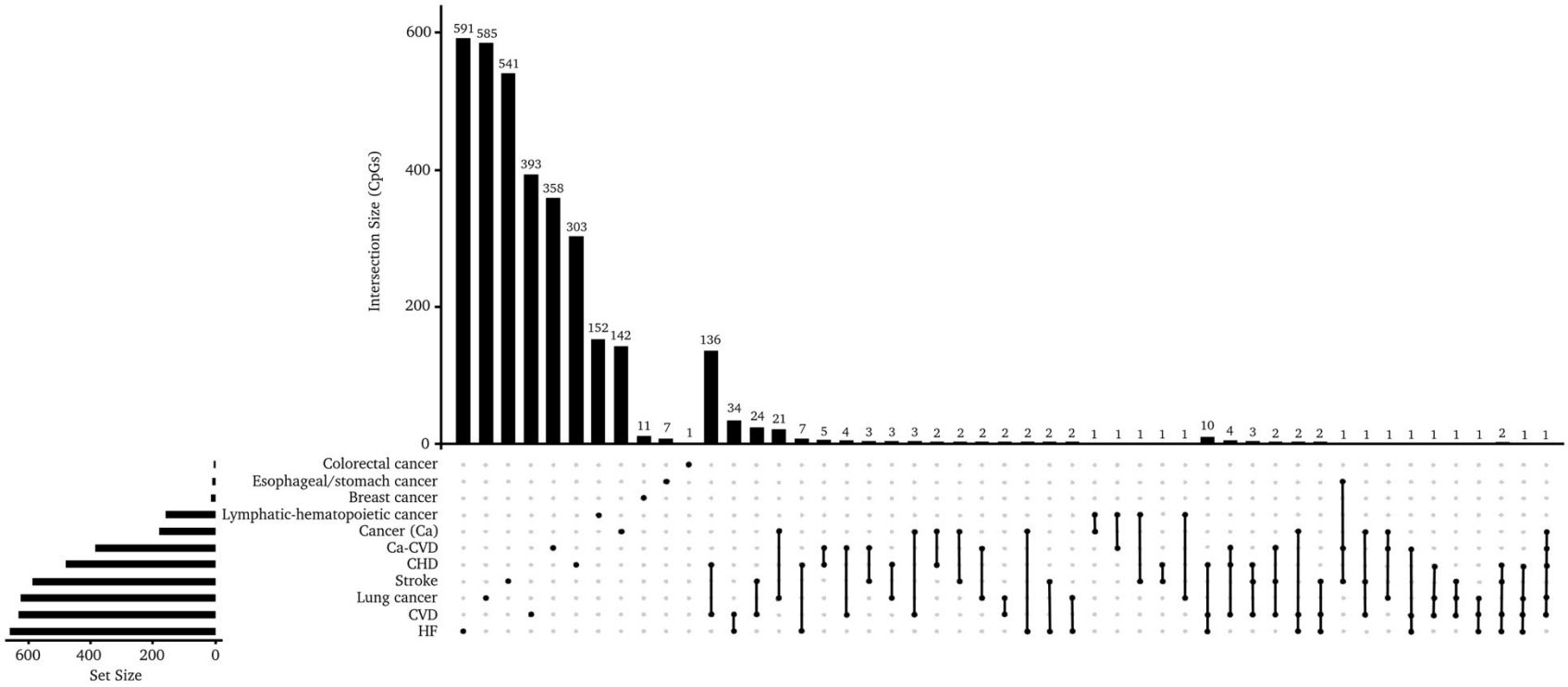


Figure S1. Up-Set plot of the intersection between all-cause and specific CVD and cancer associated CpGs.

Table S1. Predictive ability of different sets of DMPs for cancer in individuals with previous CVD events. *In cancer-specific models, we introduced a combination of Lung, Breast, Lymphatic/hematopoietic and colorectal DMPs for All-cause cancer models.

		All Cause Cancer (N cases/non cases=127/808)	Lung Cancer (N cases/non cases =23/998)	Breast (N cases/non cases =15/560)	Lymphatic/ hematopoietic (N cases/non cases =12/1011)	Esophageal/stomach (N cases/non cases =11/1011)
Models	N CpGs	C index	C index	C index	C index	C index
Reference: fully adjusted model without DMPs	0	0.63	0.85	0.73	0.84	0.50
Reference + Ca DMPs	178	0.66	0.98	0.50	0.98	0.50
Reference + Ca-CVD DMPs	383	0.96	0.99	0.99	0.96	0.99
Reference + DMPs annotated to overlapping genes for any combination of Ca, CVD and Ca-CVD DMPs (green nodes in the protein interactions network)	80	0.85	0.96	0.95	0.90	0.96

Table S2. GO terms in common for cancer and CVD. N refers to the number of the group members and n is the number in our data. The "-" symbol indicates that the GO term was not reported as significant in the group of patients.

GO ID	GO term	N	n	CVD (p-val)	Cancer (p-val)	Ca-CVD (p-val)
GO:0004611	Phosphoenolpyruvate carboxykinase activity	1	1	0.0196	-	0.0127
GO:0034344	Regulation of type III interferon production	2	1	0.0245	-	0.0211
GO:0039528	Cytoplasmic pattern recognition receptor signaling pathway in response to virus	2	1	0.0245	-	0.0211
GO:0044154	Histone H3-K14 acetylation	2	1	0.0271	-	0.0215
GO:0004613	Phosphoenolpyruvate carboxykinase (GTP) activity	2	1	0.0283	-	0.0209
GO:0043972	Histone H3-K23 acetylation	2	1	0.0360	-	0.0244
GO:0007613	Memory	74	4	0.0413	-	0.0338
GO:0039530	MDA-5 signaling pathway	3	1	0.0426	-	0.0328
GO:0004077	Biotin-[acetyl-CoA-carboxylase] ligase activity	1	1	0.0446	-	0.0275
GO:0004078	Biotin-[methylcrotonoyl-CoA-carboxylase] ligase activity	1	1	0.0446	-	0.0275
GO:0004079	Biotin-[methylmalonyl-CoA-carboxytransferase] ligase activity	1	1	0.0446	-	0.0275
GO:0004080	Biotin-[propionyl-CoA-carboxylase (ATP-hydrolyzing)] ligase activity	1	1	0.0446	-	0.0275
GO:0009305	Protein biotinylation	1	1	0.0446	-	0.0275
GO:0018271	Biotin-protein ligase activity	1	1	0.0446	-	0.0275
GO:0071110	Histone biotinylation	1	1	0.0446	-	0.0275
GO:0071034	CUT catabolic process	3	1	0.0499	-	0.0351
GO:0070888	E-box binding	49	3	-	0.0100	0.0416
GO:0035483	Gastric emptying	3	1	0.0276	0.0139	-

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, P and Q in the main text and correspond to the following results in the excel file:

A) CVD Elastic-net Cox. The 628 CpG sites selected by the elastic-net model in primary endpoints analysis for CVD group, including the traditional Cox proportional hazards model results.

B) Cancer elastic-net Cox. The 178 CpG sites selected by the elastic-net model in primary endpoints analysis for Ca group, including the traditional Cox proportional hazards model results.

C) Ca-CVD elastic-net Cox. The 383 CpG sites selected by the elastic-net model in primary endpoints analysis for Ca-CVD group, including the traditional Cox proportional hazards model results.

D) Elastic-net coronary heart disease (CHD). The 479 CpGs sites selected for CHD in the secondary endpoint elastic-net analysis.

E) Elastic-net heart failure (HF). The 657 CpGs sites selected for HF in the secondary endpoint elastic-net analysis.

F) Elastic-net stroke. The 586 CpGs sites selected for Stroke in the secondary endpoint elastic-net analysis.

G) Elastic-net lung cancer. The 622 CpGs sites selected for lung cancer in the secondary endpoint elastic-net analysis.

H) Elastic-net lymphatic-hematopoietic cancer. The 156 CpGs sites selected for lymphatic-hematopoietic cancer in the secondary endpoint elastic-net analysis.

I) Elastic-net esophageal-stomach cancer. The 8 CpGs sites selected for esophageal-stomach cancer in the secondary endpoint elastic-net analysis.

J) Elastic-net colorectal cancer. The 1 CpG site selected for colorectal cancer in the secondary endpoint elastic-net analysis.

K) Elastic-net breast cancer. The 11 CpG site selected for breast cancer in the secondary endpoint elastic-net analysis.

N) CVD GO terms. The over-represented GO terms in the CVD-DMP list. **L) Cancer GO terms.** The over-represented GO terms in the Ca-DMP list. **M) Ca-CVD GO terms.** The over-represented GO terms in the Ca-CVD-DMP list.

O) Network Nodes. The 716 nodes included in the protein interaction network (Figure 2) and its statistical parameters. **P) Network Edges.** Confidence score and other edge parameters for each of the 1622 interactions included in the network (Figure 2). **Q) Network enrichment analysis.** The subnetwork (nodes and interaction) included in the analyses and the main results.