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

SUPPLEMENTARY MATERIAL

Supplementary Methods

Cancer and CVD incidence ascertainment and follow-up definition

The Strong Heart Study (SHS): 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.[1] 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 and was reviewed independently by two or more physicians. Cancer incidence was assessed by interviews, death certificates and/or chart reviews.

The Framingham Heart Study (FHS): 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.[2] 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. Cancer incidence was assessed by interviews, death certificates, and/or chart reviews that included pathology reports, and crosschecked with official medical records whenever possible. Cancer cases included all cases of primary cancers (malignant or in-situ). Cases described in pathology reports as benign tumors or tumors of borderline malignancy and basal cell and squamous cell skin cancer cases were excluded.[3]

Atherosclerosis Risk in Communities Study (ARIC): CVD incidence included incident heart failure (HF), myocardial infarction (MI), ischemic stroke and atrial fibrillation (AF).[4] 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. Cancer incidence was identified through several methods.[5] These included linking the cohort with state cancer registries as well as active surveillance of the cohort involved recording hospital discharge codes for all participants. Participants self-reported hospitalizations during annual follow-up calls. Any cancer related hospitalizations not initially identified through registry linkage were considered cases after verification through the retrieval and review of medical records.

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 array. 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 recommendations.[6] 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,[7,8] and regression on correlated probes normalization for probe type bias.[9]

Batch effects for sample row, sample plate and sample isolation time were corrected using the combat package.[10] Blood cell counts were calculated using the Flowsorted.Blood.EPIC R package,[11] 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.[12] We annotated CpGs to the nearest gene (reference genome hg19) using the matchGenes function in *minfi* R package.[13]

The Framingham Heart Study. 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.[14] DASEN methodology in *wateRmelon* R package[15] 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 Study. Genome-wide DNAm 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.[16] In addition, 1,104 European American participants from the same cohort were profiled at 482,815 CpG sites.[17] Given that methylation data in the African American and European American 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 excluded for downstream analysis.[16,17] 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 (β) 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[16] provided a comprehensive assessment of the DNAm measurement, including the evaluation of technical replicates.

Covariate assessment

Detailed information about baseline examination, physical exam procedures and reporting of clinical covariates have been published for each cohort.[18–20] Self-report was used to ascertain sociodemographic characteristics and previous medical history. Smoking status was determined based on self-reported questionnaires. Plasma cholesterol (total, HDL and LDL cholesterol) were measured in fasting samples by enzymatic methods in unit of mg/dL. Type 2 diabetes (fasting glucose ≥7 mmol/L or antidiabetic treatment) was diagnosed by 1997 American Diabetes Association recommendations. Hypertension was defined as systolic blood pressure ≥140 mm Hg, diastolic blood pressure ≥ 90 mm Hg, or current use of antihypertensive medication. BMI was

derived from height and weight information, measured at baseline examination, as weight in kilograms (kg) divided by height in metres squared (m^2). In the Strong Heart Study, urinary albuminuria was estimated by the urinary albumin/creatinine ratio (ACR), and was modeled using pre-defined categories (<30 mg/g, 30-300 mg/g, and ≥300mg/g).

Cox elastic-net. The algorithm fits a Cox model with a penalty controlled by the λ and α parameters. λ was selected via 10-fold cross-validation to minimize the error using the cv.glmnet R function. The α parameter can range from 0 – corresponding to Ridge regression to 1 – corresponding to LASSO regression. LASSO regression is not able to deal with multicollinearity, as it only selects one variable from a correlated set. Thus, based on this and on the α choice for previous EWAS studies in the literature,[21] we selected α=0.05.All cohorts measured DNAm using the 450K microarray except the SHS, which used the EPIC. We therefore restricted analyses to CpGs included in the 450K array.

Molecular pathway analyses. We conducted a protein-protein interaction analysis of DMPs annotated to protein-coding genes identified in the second stage of the two-stage approach described in the previous subsection (targeted EWAS DMPs, Central Illustration). We obtained reported biological interactions between the protein nodes from the STRING database v11.5.[22] 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. We included database annotated interactions, automated text mining and experimental studies with a confidence score of 0.5 or greater. For exploratory analyses, we also ran an unrestricted network (confidence score of 0.0 or greater). The network was analyzed and displayed using the yFiles organic layout with Cytoscape v3.9.1.[23] A network enrichment analysis was performed by incorporating available information for the relationships between nodes based on Gene Ontology (GO) and UniProt databases. We also attempted to identify common relevant

biological mechanisms for cancer and CVD by evaluating which nodes are classified as drug targets within the DrugBank database v5.1. DrugBank combines detailed data annotations with comprehensive drug target information,[24] which enables assessment of whether gene-products are potential drug targets of specific compounds. Among the DrugBank entries, we searched for those related to the common nodes from overlapping cancer, CVD and Ca-CVD DMPs and those nodes with > 5 interactions.

REFERENCES

1. Lee ET, Welty TK, Fabsitz R, et al. The Strong Heart Study. A study of cardiovascular disease in American Indians: design and methods. *Am J Epidemiol*. 1990;132:1141–55.

2. D'Agostino RB, Vasan RS, Pencina MJ, et al. General cardiovascular risk profile for use in primary care: the Framingham Heart Study. *Circulation*. 2008;117:743–753.

3. BE K, GL S, A S. The Cancer Experience in the Framingham Heart Study Cohort. *Cancer*. 1991;67.

4. Almuwaqqat Z, Jokhadar M, Norby FL, et al. Association of Antidepressant Medication Type With the Incidence of Cardiovascular Disease in the ARIC Study. *Journal of the American Heart Association: Cardiovascular and Cerebrovascular Disease*. 2019;8.

5. Joshu CE, Barber JR, Coresh J, et al. Enhancing the Infrastructure of the Atherosclerosis Risk in Communities (ARIC) Study for Cancer Epidemiology Research: ARIC Cancer. *Cancer Epidemiol Biomarkers Prev*. 2018;27:295.

6. Domingo-Relloso A, Riffo-Campos AL, Haack K, et al. Cadmium, Smoking, and Human Blood DNA Methylation Profiles in Adults from the Strong Heart Study. *Environ Health Perspect*. 2020;128(6):67005.

7. Fortin J-P, Triche TJ, Hansen KD. Preprocessing, normalization and integration of the Illumina HumanMethylationEPIC array with minfi. *Bioinformatics*. 2016;33:btw691.

8. 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:e90–e90.

9. Niu L, Xu Z, Taylor JA. RCP: a novel probe design bias correction method for Illumina Methylation BeadChip. *Bioinformatics*. 2016;32:2659–63.

10. 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;28:882–883.

11. Salas LA, Koestler DC, Butler RA, et al. An optimized library for reference-based deconvolution of whole-blood biospecimens assayed using the Illumina HumanMethylationEPIC BeadArray. *Genome Biol*. 2018;19,64.

12. 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. *Genom Data*. 2016;9:22–4.

13. Doi A, Park I-H, Wen B, 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*. 2009;41:1350–3.

14. Joehanes R, Just AC, Marioni RE, et al. Epigenetic Signatures of Cigarette Smoking. *Circ Cardiovasc Genet*. 2016;9:436–447.
15. 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*. 2013;14:1–10.
16. Bose M, Wu C, Pankow JS, et al. Evaluation of microarray-based DNA methylation measurement using technical replicates: The atherosclerosis risk in communities (ARIC) study. *BMC Bioinformatics*. 2014;15:1–10.
17. Huan T, Joehanes R, Song C, et al. Genome-wide identification of DNA methylation QTLs in whole blood highlights pathways for cardiovascular disease. *Nature Communications 2019 10;1*:1–14.
18. Lee ET, Welty TK, Fabsitz R, et al. The Strong Heart Study. A study of cardiovascular disease in American Indians: design and methods. *Am J Epidemiol*. 1990;132:1141–55.
19. Almuwaqqat Z, Jokhadar M, Norby FL, et al. Association of Antidepressant Medication Type With the Incidence of Cardiovascular Disease in the ARIC Study. *Journal of the American Heart Association: Cardiovascular and Cerebrovascular Disease*. 2019;8(11):e012503.
20. Tsao CW, Vasan RS. Cohort Profile: The Framingham Heart Study (FHS): overview of milestones in cardiovascular epidemiology. *Int J Epidemiol*. 2015;44:1800–1813.
21. 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*. 2017;9:753–768.
22. Szklarczyk D, Gable AL, Lyon D, et al. STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res*. 2019;47:D607–D613.
23. Shannon P, Markiel A, Ozier O, et al. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Res*. 2003;13:2498–2504.
24. Knox C, Law V, Jewison T, et al. DrugBank 3.0: a comprehensive resource for “omics” research on drugs. *Nucleic Acids Res*. 2011;39.

Supplementary Table S1. Numbers of specific cancer and cardiovascular disease cases in each cohort.

	Cardiovascular disease	Cancer
Strong Heart Study	609 coronary heart disease, 291 congestive heart failure, 182 stroke cases	62 lung, 34 breast, 26 colorectal, 26 lymphatic-hematopoietic, 16 pancreatic, 15 liver, 14 ovarian, 13 kidney, 13 gallbladder, 10 esophagus-stomach, 2 endometrial, 1 thyroid cancer cases
Framingham Heart Study	70 coronary heart disease, 51 congestive heart failure cases	133 skin, 33 breast, 30 prostate, 25 lung, 10 lymphatic-hematopoietic, 10 bladder, 9 colorectal, 9 ovarian-uterus, 7 brain, 5 esophagus-stomach, 4 pancreas, 4 thyroid, 4 lip-oral cavity, 3 liver, 2 penis, one kidney, one gallbladder cancer cases
Atherosclerosis Risk in Communities (AA)	269 stroke, 731 myocardial infarction/fatal coronary heart disease, cardiac procedure, or heart failures	165 prostate, 97 breast, 77 lung, 59 colon, 45 hematopoietic and lymphatic, 27 pancreatic, 21 kidney, 15 stomach, 15 endometrial, 13 bladder, 11 brain, 9 rectal, 7 liver, 6 ovarian, 6 head/neck, 6 cervical, 3 thyroid, 2 skin cancer cases
Atherosclerosis Risk in Communities (EA)	70 stroke, 276 myocardial infarction/fatal coronary heart disease, cardiac procedure, or heart failures	52 prostate, 35 breast, 30 hematopoietic and lymphatic, 29 lung, 24 colon, 16 bladder, 16 skin, 13 endometrial, 12 brain, 7 kidney, 3 stomach, 3 ovarian, 3 rectal, 2 liver, 1 pancreatic, 1 head/neck, 1 cervical, 1 thyroid cancer cases

Abbreviations: AA, African American; EA, European American.

Supplementary Table S2. 10th and 90th percentiles of DNA methylation beta values at differentially methylated positions that were statistically significant in the meta-analysis from Table 2.

	Strong Heart Study		Framingham Heart Study		Atherosclerosis Risk in Communities - White		Atherosclerosis Risk in Communities - Black	
Percentiles	p10	p90	p10	p90	p10	p90	p10	p90
CVD								
cg27153400	0.013	0.029	0.027	0.093	0.013	0.12	0.010	0.029
cg05647197	0.93	0.96	0.77	0.86	0.90	0.98	0.82	0.98
cg01414728	0.39	0.54	0.38	0.49	0.38	0.52	0.42	0.55
Cancer-CVD								
cg22353329	0.04	0.11	0.23	0.31	0.17	0.26	0.18	0.26

Abbreviations: p90, 90th percentile; p10, 10th percentile; CVD, cardiovascular disease.

Supplementary Table S3. Genes classified as drug targets in the DrugBank database. We searched drug targets associated to the 32 overlapping nodes between at least two endpoints in 3 or more cohorts and the most connected nodes (degree > 5) in the network.

Drug ID	Drug name	Description and/or indication	Target (UniProt ID)	Endpoint (cohorts)
DB02216	S-Methylcysteine	Not Available	MGMT (P16455)	CVD (ARICb, ARICw, FHS)
DB04531	S-Benzylcysteine	Not Available		
DB00151	Cysteine	Amino acid commonly found as a component of total parenteral nutrition and used as an antidote for acetaminophen overdose. Also has antioxidant properties and is an important source of sulfur in human metabolism, and although it is classified as a non-essential amino acid, cysteine may be essential for infants, the elderly, and individuals with certain metabolic disease or who suffer from malabsorption syndromes.		
DB01593	Zinc	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.		
DB14487	Zinc acetate			
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		
DB11831	Dinitrochlorobenzene	Dinitrochlorobenzene has been used in trials studying the treatment of HIV Infections.		
DB07879	N-hydroxy-5-[(3-phenyl-5,6-dihydroimidazo[1,2-a]pyrazin-7(8H)-yl)carbonyl]thiophene-2-carboxamide	Not Available	HDAC4 (P56524)	CVD (ARICb, ARICw, SHS)
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-dio	Not Available		
DB06176	Romidepsin	Romidepsin is a histone deacetylase (HDAC) inhibitor used to treat cutaneous T-cell lymphoma.		
DB01593	Zinc	Nutritional supplement, see detailed description in this table above.		
DB14487	Zinc acetate			
DB14533	Zinc chloride			
DB14548	Zinc sulfate, unspecified form			
DB00721	Procaine	A local anesthetic of the ester type that has a slow onset and a short duration of action. It is mainly used for	DNMT3A (Q9Y6K1)	Ca-CVD (ARICb,

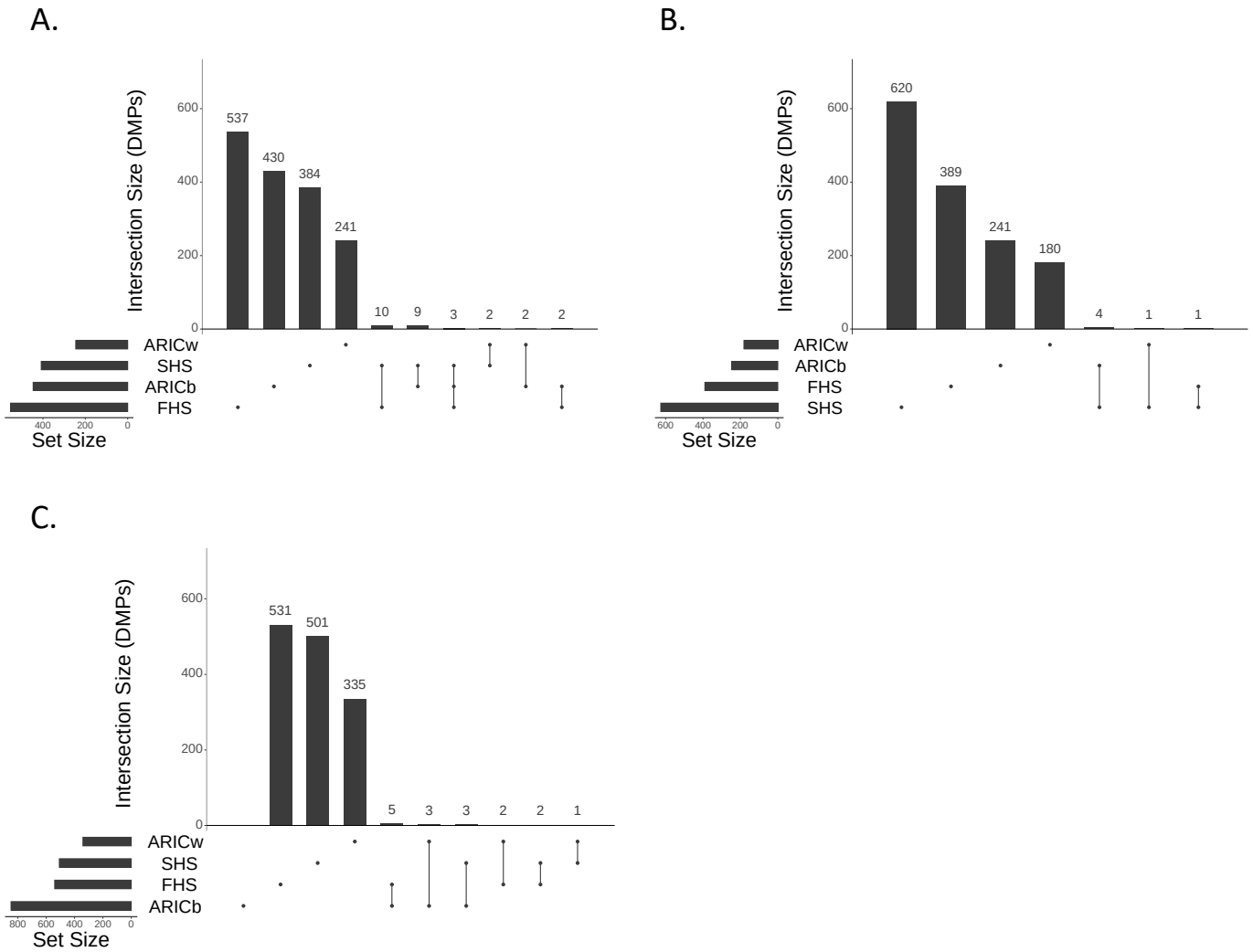
		infiltration anesthesia, peripheral nerve block, and spinal block. Procaine has also been investigated as an oral entry inhibitor in treatment-experienced HIV patients.		ARICw, FHS)
DB01262	Decitabine	A chemotherapeutic pyrimidine nucleoside analogue used for the treatment of myelodysplastic syndromes (MDS) by inducing DNA hypomethylation and corresponding alterations in gene expression.		
DB01262	Decitabine	A chemotherapeutic pyrimidine nucleoside analogue used for the treatment of myelodysplastic syndromes (MDS) by inducing DNA hypomethylation and corresponding alterations in gene expression.	DNMT3B (Q9UBC3)	Ca-CVD (ARICb, ARICw, FHS)
DB00002	Cetuximab	An endothelial growth factor receptor binding fragment used to treat colorectal cancer as well as squamous cell carcinoma of the head and neck.	EGFR (P00533)	CVD (ARICw, FHS, SHS)
DB00317	Gefitinib	A tyrosine kinase inhibitor used as first-line therapy to treat non-small cell lung carcinoma (NSCLC) that meets certain genetic mutation criteria.		
DB00530	Erlotinib	An EGFR tyrosine kinase inhibitor used to treat certain small cell lung cancers or advanced metastatic pancreatic cancers.		
DB01259	Lapatinib	An antineoplastic agent and tyrosine kinase inhibitor used for the treatment of advanced or metastatic HER-positive breast cancer in patients who received prior chemotherapeutic treatments.		
DB01269	Panitumumab	A recombinant humanized monoclonal antibody used to treat EGFR-expressing, metastatic colorectal carcinoma that is refractory to fluoropyrimidine-, oxaliplatin-, and irinotecan- containing chemotherapy regimens.		
DB00281	Lidocaine	A local anesthetic used in a wide variety of superficial and invasive procedures. In doing so, however, it can block or decrease muscle contractile, resulting in effects like vasodilation, hypotension, and irregular heart rate, among others. As a result, lidocaine is also considered a class Ib anti-arrhythmic agent.		
DB03496	Alvocidib	A synthetic flavonoid based on an extract from an Indian plant for the potential treatment of cancer. It works by inhibiting cyclin-dependent kinases, arresting cell division and causing apoptosis in non-small lung cancer cells.		
DB04988	IGN311	A humanized monoclonal antibody (mab) against the Lewis Y carbohydrate antigen, a blood-group-related oligosaccharide.		
DB05294	Vandetanib	An antineoplastic kinase inhibitor used to treat symptomatic or progressive medullary thyroid cancer in patients with unresectable locally advanced or metastatic disease.		
DB05374	Rindopepimut	An injectable peptide cancer vaccine which targets a mutant protein called EGFRvIII present in about 25% to 30% of glioblastoma cases. The vaccine consists of the EGFRv3-specific peptide conjugated to the non-specific immunomodulator keyhole limpet hemocyanin (KLH).		
DB05101	Matuzumab	A humanized monoclonal antibody used in cancer treatment. It has a high affinity for EGFR (epithelial growth factor receptor), frequently associated with the		

		growth of blood vessels in malignancy, facilitating tumor growth and survival.		
DB05424	Canertinib	A pan-erbB tyrosine kinase inhibitor which work against esophageal squamous cell carcinoma in vitro and in vivo. Canertinib treatment significantly affects tumour metabolism, proliferation and hypoxia as determined by PET		
DB05944	Varlitinib	Anti-tumor activity in preclinical models of human breast, lung, and epidermal carcinoma tumors.		
DB06021	AV-412	Not Available		
DB05524	Pelitinib	A potent, low molecular weight, selective, and irreversible inhibitor of epidermal growth factor receptor (EGFR) that is being developed as an anticancer agent.		
DB07602	S-{3-[(4-anilinoquinazolin-6-YL)amino]-3-oxopropyl}-L-cysteine	Not Available		
DB07662	PD-168393	An epidermal growth factor receptor inhibitor		
DB08916	Afatinib	An antineoplastic agent used for the treatment of locally advanced or metastatic non-small cell lung cancer with non-resistant EGFR mutations or resistance to platinum-based chemotherapy.		
DB09559	Necitumumab	A monoclonal antibody used to treat metastatic squamous non-small cell lung cancer.		
DB09330	Osimertinib	A tyrosine kinase inhibitor used in the treatment of certain types of non-small cell lung carcinoma.		
DB11737	Icotinib	A potent and specific epidermal growth factor receptor (EGFR) tyrosine kinase inhibitor (TKI). Phase I study for the evaluation of icotinib as a treatment of EGFR+ Non-Small Cell Lung Cancer (NSCLC).		
DB13164	Olmutinib	Orally active epidermal growth factor receptor inhibitor used in the treatment of T790M mutation positive non-small cell lung cancer. It is available under the brand name Olita made by Hanmi Pharmaceuticals 4.		
DB12202	Zalutumumab	Fully human IgG1 monoclonal antibody designed to bind with selectivity to the epidermal growth factor receptor (EGFR). Zalutumumab has been investigated for the treatment of Squamous Cell Cancer and Head and Neck Cancer.		
DB11731	Depatuxizumab mafodotin	Has been used in trials studying the treatment of Lymphoma, Gliosarcoma, Glioblastoma, Malignant Glioma, Squamous Cell Tumors, and Glioblastoma Multiforme.		
DB11828	Neratinib	A protein kinase inhibitor used to treat breast cancer that over expresses the HER2 receptor.		
DB12267	Brigatinib	An anaplastic lymphoma kinase inhibitor used to treat anaplastic lymphoma kinase positive metastatic non small cell lung cancer.		
DB10772	Foreskin keratinocyte (neonatal)	A treatment of skin cells used to accelerate wound closure and healing.		

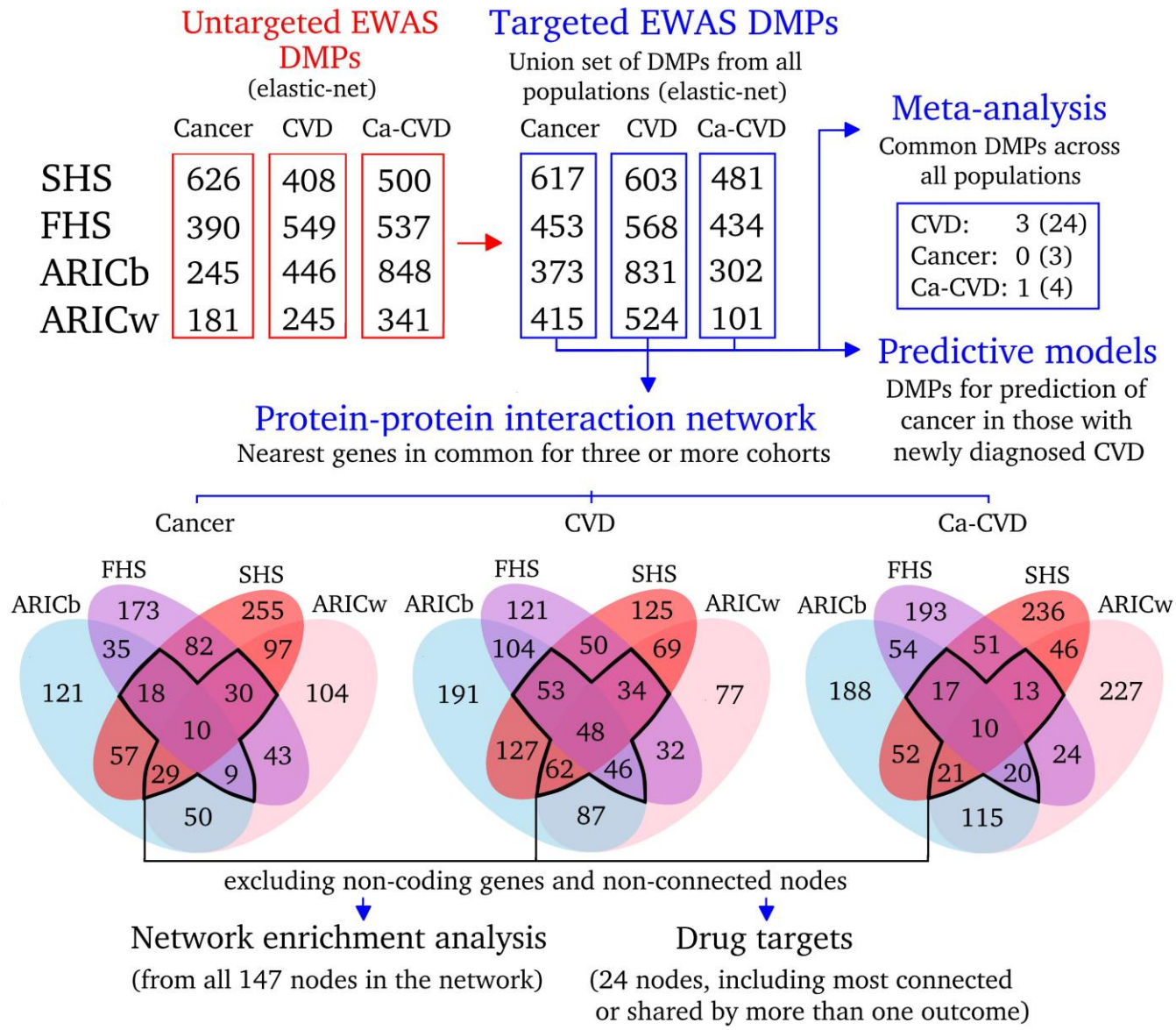
DB12010	Fostamatinib	A spleen tyrosine kinase inhibitor used to treat chronic immune thrombocytopenia after attempting one other treatment.		
DB11963	Dacomitinib	A medication used to treat non small cell lung cancer with EGFR exon 19 deletion of exon 21 L858R substitution.		
DB15035	Zanubrutinib	A kinase inhibitor used to treat mantle cell lymphoma, a type of B-cell non-Hodgkin lymphoma, in adults who previously received therapy.		
DB15327	Abivertinib	A tyrosine kinase inhibitor targeted against mutant forms of both human epidermal growth factor receptor (EGFR) and Bruton's tyrosine kinase (BTK). It has been investigated for use in the treatment of non-small cell lung cancer (NSCLC) and B-cell malignancies. Abivertinib's potential to depress cytokine production has led to its investigation in the treatment of hospitalized patients with moderate-to-severe COVID-19		
DB16695	Amivantamab	An EGF and MET receptor targeted antibody indicated in the treatment of non-small cell lung cancer with an EGFR 20 exon insertion mutation.		
DB16390	Mobocertinib	An oral kinase inhibitor targeted against EGFR and used in the treatment of NSCLC with EGFR exon 20 insertion mutations.		
DB00530	Erlotinib	An EGFR tyrosine kinase inhibitor used to treat certain small cell lung cancers or advanced metastatic pancreatic cancers.		
DB12114	Poziotinib	As been used in trials studying the treatment of Breast Cancer, Metastatic Breast Cancer, Increased Drug Resistance, Adenocarcinoma of Lung Stage IV, and Adenocarcinoma of Lung Stage IIIB, among others.		
DB02587	Colforsin	Potent activator of the adenylate cyclase system and the biosynthesis of cyclic AMP. From the plant Coleus forskohlii. Has antihypertensive, positive inotropic, platelet aggregation inhibitory, and smooth muscle relaxant activities; also lowers intraocular pressure and promotes release of hormones from the pituitary gland.	ADCY3 (O60266)	CVD (ARICb, ARICw, FHS, SHS)
DB06843	2',5'-dideoxy-adenosine 3'-monophosphate	Not Available	GNAS (P63092)	CVD (ARICb, ARICw, SHS)
DB02587	Colforsin	Potent activator of the adenylate cyclase system and the biosynthesis of cyclic AMP. From the plant Coleus forskohlii. Has antihypertensive, positive inotropic, platelet aggregation inhibitory, and smooth muscle relaxant activities; also lowers intraocular pressure and promotes release of hormones from the pituitary gland.		
DB01812	Adenosine 3',5'-diphosphate	Not Available	HS3ST3A1 (Q9Y663)	Cancer (ARICb, FHS, SHS); CVD (ARICb,
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		ARICw, SHS)

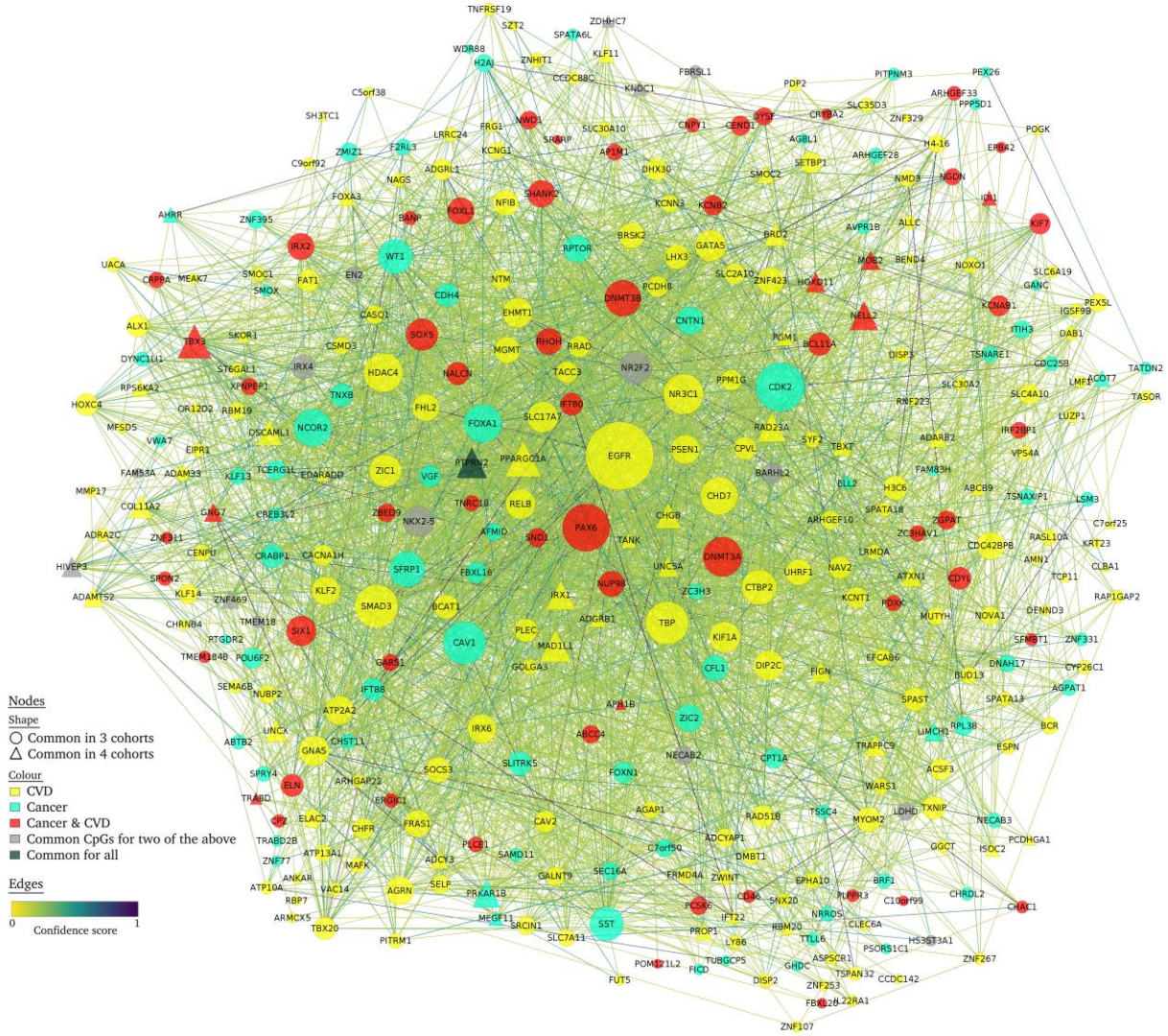
Supplementary Figure 1. Upset plot of the intersection of DMPs across cohorts in the untargeted EWAS for A) CVD B) Cancer C) Cancer-CVD.



Supplementary Figure 2. Workflow of the multi-cohort study (extension of the Central Illustration).



Supplementary Figure 3. Unrestricted network.



The network was analyzed by Cytoscape and included 352 nodes and 3734 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 0.0 or greater, were obtained from the STRING database.

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) Untargeted CVD DMPs in the SHS. CpG sites selected by the elastic-net model in the untargeted approach for CVD.

B) Untargeted cancer DMPs in the SHS. CpG sites selected by the elastic-net model in the untargeted approach for cancer.

C) Untargeted Ca-CVD DMPs in the SHS. CpG sites selected by the elastic-net model in the untargeted approach for Ca-CVD.

D) Untargeted CVD DMPs in the FHS. CpG sites selected by the elastic-net model in the untargeted approach for CVD.

E) Untargeted cancer DMPs in the FHS. The CpG sites selected by the elastic-net model in the untargeted approach for cancer.

F) Untargeted Ca-CVD DMPs in the FHS. CpG sites selected by the elastic-net model in the untargeted approach for Ca-CVD.

G) Untargeted CVD DMPs in ARIC African Americans. CpG sites selected by the elastic-net model in the untargeted approach for CVD.

H) Untargeted cancer DMPs in ARIC African Americans. CpG sites selected by the elastic-net model in the untargeted approach for cancer.

I) Untargeted Ca-CVD DMPs in ARIC African Americans. CpG sites selected by the elastic-net model in the untargeted approach for Ca-CVD.

J) Untargeted CVD DMPs in ARIC European Americans. CpG sites selected by the elastic-net model in the untargeted approach for CVD.

K) Untargeted cancer DMPs in ARIC European Americans. CpG sites selected by the elastic-net model in the untargeted approach for cancer.

L) Untargeted Ca-CVD DMPs in ARIC European Americans. CpG sites selected by the elastic-net model in the untargeted approach for Ca-CVD.

M) Network Nodes. Nodes included in the protein interaction network (Figure 1) and corresponding statistical parameters.

N) Network Edges. Confidence score and other edge parameters for the interactions included in the network (Figure 1).

O) Network enrichment analysis. All significantly enriched terms that STRING return for the 147 nodes of the network.

P) DrugBank database target search. Overlapping nodes/genes in cancer, CVD and Ca-CVD DMPs. Accessed March 17, 2023.

Q) Supplementary Unrestricted Network Nodes. Nodes included in the protein interaction unrestricted network (Supplementary Figure 3) and corresponding statistical parameters.

R) Supplementary Unrestricted Network Edges. Confidence score and other edge parameters for the interactions included in the unrestricted network (Supplementary Figure 3).

S) Supplementary Unrestricted Network enrichment analysis. All significantly enriched terms that STRING return for the 358 nodes of the network.