ARIC Manuscript Proposal #3864

PC Reviewed: 6/8/21	**Status: _____**	**Priority: 2**
SC Reviewed: _____	**Status: _____**	**Priority: ____**

1.a. Full Title: Epigenome-wide association study of all-cause cancer and cardiovascular disease incidence: Common blood DNA methylation signatures

b. Abbreviated Title (Length 26 characters): CVD, cancer and DNA methylation

2. Writing Group:

Writing group members:

Angela L. Riffo-Campos
Arce Domingo-Relloso
Guillermo Ayala
Karin Haack
Andrés Ávila
Dorothy A. Rhoades
Jason G. Umans
M Daniele Fallin
Miguel Herreros-Martinez
Marina Pollan
Shelley A. Cole
Elizabeth Platz
Miranda Jones
Dominique Michaud
Jan Bressler
Ana Navas-Acien
Maria Tellez-Plaza

I, the first author, confirm that all the coauthors have given their approval for this manuscript proposal. _ALRC_ **[please confirm with your initials electronically or in writing]**

First author: **Angela L Riffo-Campos**
Address: Centro de Excelencia de Modelación y Computación Científica
Universidad de La Frontera
Montevideo Street, 0785
4780000 Temuco-Chile
Phone: Fax:
E-mail: angela.riffo@ufrontera.cl

ARIC author to be contacted if there are questions about the manuscript and the first author does not respond or cannot be located (this must be an ARIC investigator).

Name: **Elizabeth Platz**

Address: Department of Epidemiology
Johns Hopkins Bloomberg School of Public Health
615 N. Wolfe Street
Room E6132
Baltimore, Maryland 21205
Phone: 410-614-9674 Fax: (410) 614-2632
E-mail: eplatz1@jhu.edu

3. Timeline:

We expect to submit this manuscript for publication within four months of approval.

4. Rationale:

Cardiovascular disease (CVD) and cancer are the most common causes of death. Both diseases share risk factors including smoking, physical inactivity, unhealthy diet, excess adiposity, genetic predisposition, environmental toxicants, advanced age and others (1–4). Emerging evidence reveals a complex connection between CVD and cancer (5,6), as they share inter-related biological pathways, including chronic inflammation, oxidative stress, aberrant apoptosis and angiogenesis (2,3,7). Cardiotoxicity can follow cancer treatment, such as chemotherapy, radiotherapy (3,4,8,9) and, to a lesser extent, targeted cancer therapies and immunotherapy (10). A deeper understanding of the molecular relationships between CVD and cancer can help identify new and safer preventive and therapeutic approaches for both diseases.

Epigenetic modifications regulate gene expression (11) and are involved in biological functions that influence health and disease (12). Cancer and CVD epigenetic changes, however, are traditionally studied separately (13,14). For instance, DNA methylation (DNAm), the addition of a methyl group in the carbon-5 position of cytosine followed by guanine separated by a phosphate (CpG) (12) has not yet been jointly evaluated for cancer and CVD outcomes.

We will identify common epigenetic modifications for CVD and cancer by studying the overlap of differentially methylated positions (DMPs) associated with incident cardiovascular disease (CVD), cancer (Ca), and both (Ca-CVD) from separate epigenome-wide association studies (EWAS). The study will be conducted in the Strong Heart Study (SHS), 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 as well as in the Atherosclerosis Risk in Communities (ARIC) Study and the Framingham Heart Study (FHS). In each cohort, we will jointly model a panel of ~450000 CpGs measured by the Illumina Infinium HumanMethylation450 BeadChip array (in the SHS the Infinium MethylationEPIC BeadChip is available, but we will restrict the data to the Illumina 450K as these are the CpGs available in the other cohorts, including ARIC). Using the CpGs identified as relevant for the 3 outcomes in each cohort, we will further evaluate those CpGs in the other cohorts. We have successfully used this approach in a study of DNAm and coronary heart disease that is currently in press at *JAMA Cardiology*. We will subsequently conduct molecular pathway analyses and other integrative analysis of findings. The purpose of this paper proposal is to request approval to apply this methodology in ARIC. The ARIC study is a unique setting to evaluate the epigenetic relationship between cancer and CVDs

and to compare these signals to SHS and FHS. It has 14,688 participants, including cancer incidence cases and over 30 years of follow-up for CVD.

5. Main Hypothesis/Study Questions:

Hypothesis: Some of the common DMPs between cardiovascular disease and cancer identified in SHS will also be identified in the ARIC study and the FHS, and reversely, some of the common DMPs between cardiovascular disease and cancer identified in ARIC will also be identified as relevant in the SHS and FHS, using the same methodology.

The main objective is to identify DMPs that are shared between CVD and cancer, and which are the molecular pathways and networks shared across outcomes and commonly identified across cohorts. This analysis can help us identify relevant DMPs for cancer and CVD. While many pathways are known and might just be replicated in this analyses, other novel connections might be identified in this study.

Aim 1: To study the epigenetic relationship between cancer and cardiovascular disease in the ARIC study using four groups of participants: incident cardiovascular disease (CVD), cancer (Ca), both (Ca-CVD) and non-cases (common reference).

Aim 2: To compare the DMPs identified in ARIC with those obtained from the SHS and FHS.

Aim 3: To conduct an integrative *In-silico* protein-protein interaction analysis to identify most relevant biological pathways that could be intervened in future research and might help the identification of potential precision medicine tools.

6. Design and analysis (study design, inclusion/exclusion, outcome and other variables of interest with specific reference to the time of their collection, summary of data analysis, and any anticipated methodologic limitations or challenges if present).

Study design: prospective cohort study

Three different outcomes: incident cardiovascular disease (CVD) in those who do not develop cancer over the follow-up; cancer (Ca) in those who do not develop CVD over the follow-up; and a combined outcome for those who develop both incident CVD and cancer during the follow-up (in this group the time to event will be the date of the first diagnosis, either CVD or cancer, which ever occurs first).

Inclusions and Exclusions
Individuals with prevalent CVD or cancer should be excluded for all analyses.
In cardiovascular disease only: Both fatal and non-fatal all-cause CVD events should be considered. Individuals who developed cancer at any time of the follow-up period should be excluded from this analysis. CVD events include coronary heart disease, stroke, and heart failure.

In cancer only: Both fatal and non-fatal all-cause cancer events should be considered. Individuals who developed CVD at any time of the follow-up period should be excluded from this analysis.
In CVD & cancer: only individuals who developed both CVD and cancer during the follow-up period should be considered. Individuals who developed CVD but not cancer and those who developed cancer but not CVD should be excluded.
In Non cases: for all three groups of cases, non-cases are those who never developed CVD or cancer during the follow-up.

DNA methylation data preparation
Preprocessing and quality control is determined by the individual analysts of each study. For example, the Strong Heart Study excluded CpGs with a p-detection value greater than 0.01 in more than 5% of the individuals and individuals with low median intensities (<10). Single sample snoob normalization was conducted using the R package minfi (15,16). Cross-hybridizing probes, sex chromosomes and SNP probes with minor allele frequency > 0.05 (17,18) were removed for analysis. Houseman cell proportions (CD8T, CD4T, NK, B cells, monocytes and granulocytes) were estimated using the R package FlowSorted.Blood.EPIC (Houseman method). Batch effects were detected and corrected using the combat function (sva R package). We will follow the steps already established by ARIC.

Statistical analysis
Descriptive analysis. A descriptive table of the participant will be prepared, similar to this example:

	Non cases	Incident cancer (no CVD) cases	Incident CVD (no cancer) cases	Incident Ca-CVD cases
Number of participants				
Sum of person-years				
Age				
Female %				
Never smoking %				
Former smoking %				
Current smoking %				
Packyears smoked				
Methylation-predicted packyears smoked				
BMI kg/m^2				
Diabetes %				
LDL-cholesterol, mg/dL				
HDL-cholesterol, mg/dL				
Hypertension %				

SBP, mmHg
Albuminuria > 30 mg/g, % *
Aspirin
Lipid-lowering drugs
Female HRT use

* Only if data on albuminuria status is available

Elastic-net models. Each cohort will run three epigenome-wide elastic-net models (one for each outcome) using the R package glmnet. The alpha value will be set to 0.05 and the lambda value will be selected by cross-validation (R code will be provided). C indexes for each elastic-net model will be calculated.

Cancer models will be adjusted for the following covariates: age, sex, BMI, smoking status, blood cell counts, methylation-predicted packyears smoked (19) and technical covariates (specific to each population).

CVD models will be additionally adjusted for the following covariates: LDL cholesterol, hypertension, type 2 diabetes, systolic blood pressure, HDL cholesterol and, if available, albuminuria status.

The CVD adjustment set will be taken in all the models, including the combined outcome of cancer and CVD.

For each of the outcomes, the cohorts will run an additional elastic-net model including the union set of CpGs that have been selected by elastic-net by each of the cohorts. Then, traditional Cox regression models will be run for the CpGs commonly selected by the cohorts. A meta-analysis for the hazard ratios will be conducted (for each of the outcomes separately) in CpGs that are common across cohorts in this second elastic-net model. As power might be limited in some cohorts, we will also run the meta-analysis for CpGs that are selected in at least 3 of the 4 cohorts.

Molecular pathways analyses. The DMPs associated to protein coding genes will be included in a protein-protein interaction network, after obtaining reported biological interactions between the protein nodes from the STRING database v11.0 (20). All interactions with a confidence score of 0.5 or greater will be included (Figure S1). The networks will be analyzed and displayed using the yFiles organic layout with Cytoscape v3.7.1 (21). A network enrichment analysis will be performed by incorporating available information for the relationships between nodes based on GO, KEGG, Reactome and UniProt databases through a combination of methods that include multiple testing correction, two-sided Kolmogorov–Smirnov test, and hierarchical clustering of the STRING network itself (21). We also attempted to identify common relevant biological mechanisms for Ca and CVD by evaluating which nodes are drug targets within DrugBank database v5.1.7. DrugBank combines detailed data annotations with comprehensive drug target information (22), which enables assessing if gene-products are potential drug targets of specific targeting compounds. Among the 5,225 non-redundant proteins corresponding to drug entries and associated drug targets in DrugBank, we will search for those related to the common nodes from overlapping Ca, CVD and Ca-CVD DMPs.

Limitations:

Even though the sample size might be moderate, we already conducted a DMP analysis for coronary heart disease in ARIC black (N=2,114; 350 cases), ARIC white (N=931, 121 cases) and the SHS (N=2,321; 749 cases) for the previously mentioned paper that is currently in press at *JAMA Cardiology*, which suggests we have sufficient power for the proposed elastic-net approach.

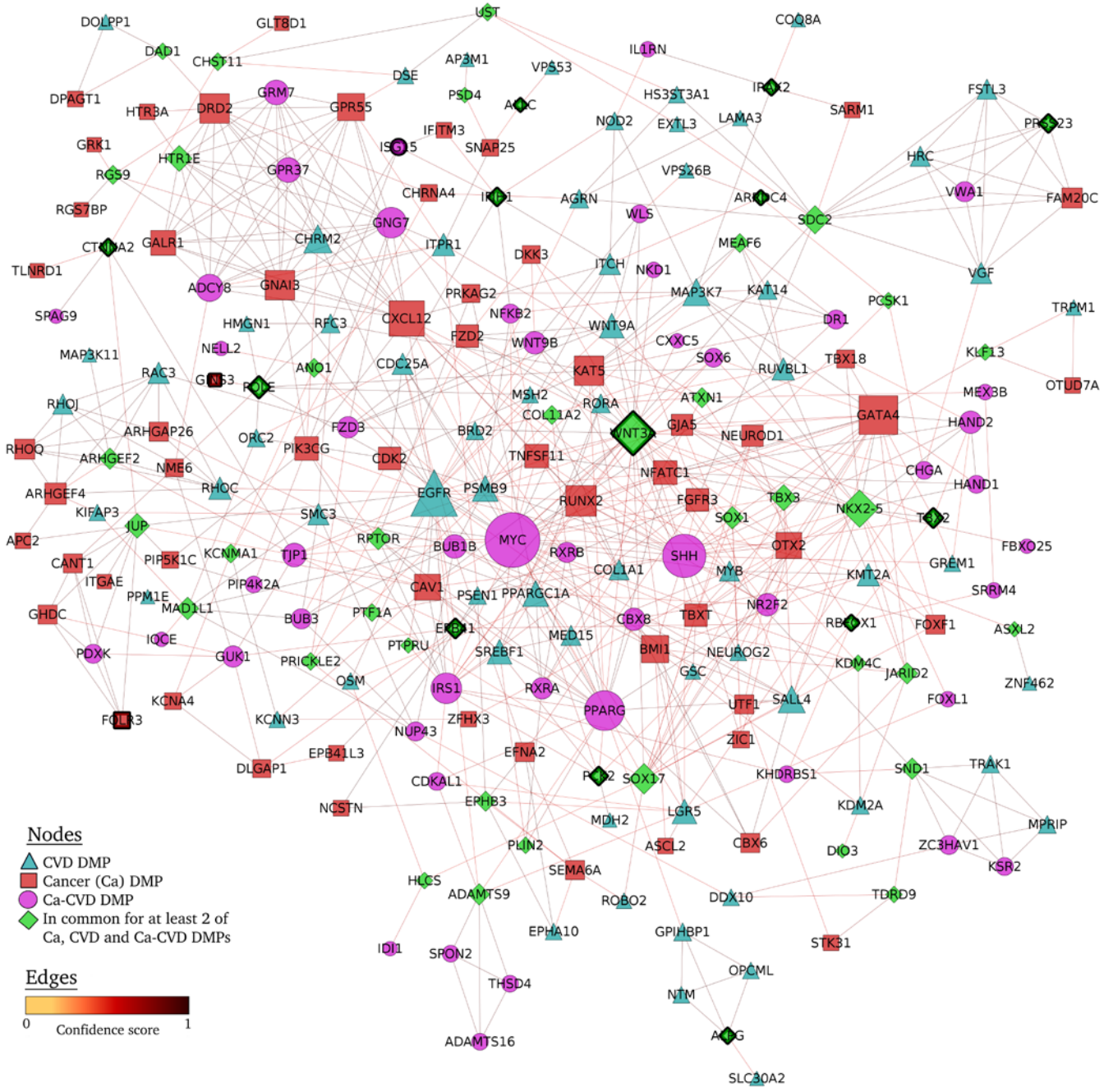
On the other hand, it is known that cancer and CVD share common risk factors. In this study, we are attempting to adjust for smoking as well as possible based on CpG signatures of smoking and pack-years in order to identify DMPs associated with cancer and CVD beyond smoking. For other risk factors, it is harder to do as signatures for those factors are still under active investigation. It is thus possible that some of the DMPs identified in this study reflect those risk factors. In part, we might be able to assess this phenomenon by observing if traditional risk factors are no longer kept by the elastic net models. We will also evaluate how the top DMPs are associated with CVD and cancer risk factors (e.g., obesity, lipids, diabetes, etc.). Also, if during the conduction of this study other epigenetic signatures are developed beyond smoking, we will also incorporate those into the analysis.

Although the prevalence of risk factors and the distribution of the specific outcomes might be different across cohorts (e.g., diabetes and liver cancer are more common in the SHS than in ARIC), we believe that our proposed strategy will still contribute to identify general CpGs that are relevant for cardiovascular and cancer outcomes. For instance, in the previously cited study of DNAm and coronary heart disease that is currently in press in *JAMA Cardiology*, there was a substantial number of common DMPs between SHS and ARIC. At the same time there were also many distinct DMPs, which could reflect the specific differences across cohorts.

Previous experience in the SHS:

Our preliminary findings in the Strong Heart Study (n=2,186; 823, 277, and 142 cases of CVD only, Cancer only, and Cancer-CVD, respectively) support there is a network of DMPs annotated to biologically interconnected genes that are related to cancer and CVD outcomes, as observed in our protein-protein interaction network (Figure 1). In our previous study focused on CHD (in press in *JAMA Cardiology*), our proposed strategy for this manuscript allowed to identify a subset of DMPs that were common across 4 or 5 cohorts, as well as DMPs that were distinct across cohorts. Given this previous experience, which used the same cohorts and had similar sample sizes, we believe we can identify a set of DMPs that are commonly related to Cancer and CVD that replicated across diverse cohorts.

Figure 1. Protein-protein interaction network for protein-coding genes associated with cancer, CVD or cancer-CVD endpoints.



7.a. Will the data be used for non-ARIC analysis or by a for-profit organization in this manuscript? ____ Yes __X__ No

b. If Yes, is the author aware that the current derived consent file ICTDER05 must be used to exclude persons with a value RES_OTH and/or RES_DNA = "ARIC only" and/or "Not for Profit" ? ____ Yes ____ No

(The file ICTDER has been distributed to ARIC PIs, and contains
the responses to consent updates related to stored sample use for research.)

8.a. Will the DNA data be used in this manuscript? __X__ Yes ____ No

8.b. If yes, is the author aware that either DNA data distributed by the Coordinating Center must be used, or the current derived consent file ICTDER05 must be used to exclude those with value RES_DNA = "No use/storage DNA"? __X__ Yes ____ No

9. The lead author of this manuscript proposal has reviewed the list of existing ARIC Study manuscript proposals and has found no overlap between this proposal and previously approved manuscript proposals either published or still in active status. ARIC Investigators have access to the publications lists under the Study Members Area of the web site at: http://www.cscc.unc.edu/aric/mantrack/maintain/search/dtSearch.html

__X__ Yes _____ No

10. What are the most related manuscript proposals in ARIC (authors are encouraged to contact lead authors of these proposals for comments on the new proposal or collaboration)?

The previous paper proposals that consider only cancer methylation (for example ARIC Manuscript Proposal #3248) or only CVD methylation (for example ARIC Manuscript Proposal #1928) have no common objectives with this proposal, since our main interest is studying the epigenetics relationships of cancer and CVD together. For this reason, the paper proposals that are most related to the subject of study are those including both CVD and cancer outcomes as mentioned below.

ARIC Manuscript Proposal #3563: Polter et al. CVD risk in Cancer Survivors
ARIC Manuscript Proposal #3257: Moran et al. Heart Failure Prediction in Cancer Survivors
ARIC Manuscript Proposal #3028: Florido et al. Cardiovascular Risk among Cancer Survivors in the ARIC study
ARIC Manuscript Proposal #2912: Florido et al. Subclinical Myocardial Damage among Cancer Survivors in ARIC

Previous Manuscript Proposals were focused on specific types of CVD developed by cancer survivor participants in ARIC. In this paper proposal, we investigate the underlying epigenetic mechanisms of the relationship between the incidence of both all cancer and all CVD, using four groups of participants: Ca, CVD, Ca-CVD and non-cases (excluding prevalent CVD or cancer). Therefore, the approach is presented here for the first time.

11.a. Is this manuscript proposal associated with any ARIC ancillary studies or use any ancillary study data? __X__ Yes ___ No

11.b. If yes, is the proposal

_X__ A. primarily the result of an ancillary study (list number* _____

Enhancing the Infrastructure of the Atherosclerosis Risk in Communities (ARIC) Study for Cancer Epidemiology Research: ARIC Cancer (PMID: 29263187)
Novel associations between blood DNA methylation and body mass index in middle-aged and older adults (PMID: 29278407)

___ B. primarily based on ARIC data with ancillary data playing a minor role (usually control variables; list number(s)* _____ _____ _____)

*ancillary studies are listed by number https://sites.cscc.unc.edu/aric/approved-ancillary-studies

12a. Manuscript preparation is expected to be completed in one to three years. If a manuscript is not submitted for ARIC review at the end of the 3-years from the date of the approval, the manuscript proposal will expire.

12b. The NIH instituted a Public Access Policy in April, 2008 which ensures that the public has access to the published results of NIH funded research. It is **your responsibility to upload manuscripts to PubMed Central** whenever the journal does not and be in compliance with this policy. Four files about the public access policy from http://publicaccess.nih.gov/ are posted in http://www.cscc.unc.edu/aric/index.php, under Publications, Policies & Forms. http://publicaccess.nih.gov/submit_process_journals.htm shows you which journals automatically upload articles to PubMed central.

References

1. Friedrich CA, Nettleton JA. Cancer : The Atherosclerosis Risk in Communities Study. Circulation. 2014;127(12):1270–5.
2. Koene RJ, Prizment AE, Blaes A, Konety SH. Shared Risk Factors in Cardiovascular Disease and Cancer. Circulation. 2016 Mar 15;133(11):1104-14. doi: 10.1161/CIRCULATIONAHA.115.020406.
3. Giza DE, Iliescu G, Hassan S, Marmagkiolis K, Iliescu C. Cancer as a Risk Factor for Cardiovascular Disease. Curr Oncol Rep. 2017 Jun;19(6):39. doi: 10.1007/s11912-017-0601-x.
4. Al-Kindi SG, Oliveira GH. Onco-Cardiology: A Tale of Interplay Between 2 Families of Diseases. Mayo Clin Proc. 2016 Dec;91(12):1675-1677. doi: 10.1016/j.mayocp.2016.10.015.
5. Masoudkabir F, Sarrafzadegan N, Gotay C, Ignaszewski A, Krahn AD, Davis MK, Franco C, Mani A. Cardiovascular disease and cancer: Evidence for shared disease pathways and pharmacologic prevention. Atherosclerosis. 2017 Aug;263:343-351. doi: 10.1016/j.atherosclerosis.
6. Wang L, Wang F, Chen L, et al. Long-term cardiovascular disease mortality among 160 834 5-year survivors of adolescent and young adult cancer: an American population-based cohort study. European Heart Journal. 2021 Jan;42(1):101-109. DOI: 10.1093/eurheartj/ehaa779.

7. Tapia-Vieyra JV, Delgado-Coello B, Mas-Oliva J. Atherosclerosis and Cancer; A Resemblance with Far-reaching Implications. Arch Med Res. 2017 Jan;48(1):12-26. doi: 10.1016/j.arcmed.2017.03.005.
8. Hong RA, Iimura T, Sumida KN, Eager RM. Cardio-oncology/onco-cardiology. Clin Cardiol. 2010 Dec;33(12):733-7. doi: 10.1002/clc.20823.
9. Aleman BM, Moser EC, Nuver J, Suter TM, Maraldo MV, Specht L, Vrieling C, Darby SC. Cardiovascular disease after cancer therapy. EJC Suppl. 2014 Jun;12(1):18-28. doi: 10.1016/j.ejcsup.2014.03.002.
10. Michel L, Schadendorf D, Rassaf T. Oncocardiology: new challenges, new opportunities. Herz. 2020 Nov;45(7):619-625. doi: 10.1007/s00059-020-04951-x.
11. Allis CD, Jenuwein T. The molecular hallmarks of epigenetic control. Nat Rev Genet. 2016 Aug;17(8):487-500. doi: 10.1038/nrg.2016.59.
12. Zoghbi HY, Beaudet AL. Epigenetics and Human Disease. Cold Spring Harb Perspect Biol. 2016 Feb 1;8(2):a019497. doi: 10.1101/cshperspect.a019497.
13. Kinnaird A, Zhao S, Wellen KE, Michelakis ED. Metabolic control of epigenetics in cancer. Nat Rev Cancer. 2016 Nov;16(11):694-707. doi: 10.1038/nrc.2016.82.
14. van der Harst P, de Windt LJ, Chambers JC. Translational Perspective on Epigenetics in Cardiovascular Disease. J Am Coll Cardiol. 2017 Aug 1;70(5):590-606. doi: 10.1016/j.jacc.2017.05.067.
15. Fortin J-P, Triche TJ, Hansen KD, Hansen KD. Preprocessing, normalization and integration of the Illumina HumanMethylationEPIC array with minfi. Bioinformatics. Oxford University Press; 2017;33:558–60.
16. 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. Available from: https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gkt090
17. McCartney DL, Walker RM, Morris SW, McIntosh AM, Porteous DJ, Evans KL. Identification of polymorphic and off-target probe binding sites on the Illumina Infinium MethylationEPIC BeadChip. Genomics data. Elsevier; 2016 [cited 2018 Jun 8];9:22–4. Available from: http://www.ncbi.nlm.nih.gov/pubmed/27330998
18. Illumina Inc. Infinium MethylationEPIC Product Files. 2019. Available from: https://support.illumina.com/downloads/infinium-methylationepic-v1-0-product-files.html
19. Joehanes R, Just AC, Marioni RE, Pilling LC, Reynolds LM, Mandaviya PR, Guan W, Xu T, Elks CE, Aslibekyan S, Moreno-Macias H, Smith JA, Brody JA, Dhingra R, Yousefi P, Pankow JS, Kunze S, Shah SH, McRae AF, Lohman K, Sha J, Absher DM, Ferrucci L, Zhao W, Demerath EW, Bressler J, Grove ML, Huan T, Liu C, Mendelson MM, Yao C, Kiel DP, Peters A, Wang-Sattler R, Visscher PM, Wray NR, Starr JM, Ding J, Rodriguez CJ, Wareham NJ, Irvin MR, Zhi D, Barrdahl M, Vineis P, Ambatipudi S, Uitterlinden AG, Hofman A, Schwartz J, Colicino E, Hou L, Vokonas PS, Hernandez DG, Singleton AB, Bandinelli S, Turner ST, Ware EB, Smith AK, Klengel T, Binder EB, Psaty BM, Taylor KD, Gharib SA, Swenson BR, Liang L, DeMeo DL, O'Connor GT, Herceg Z, Ressler KJ, Conneely KN, Sotoodehnia N, Kardia SL, Melzer D, Baccarelli AA, van Meurs JB, Romieu I, Arnett DK, Ong KK, Liu Y, Waldenberger M, Deary IJ, Fornage M, Levy D, London SJ. Epigenetic Signatures of Cigarette Smoking. Circ Cardiovasc Genet. 2016 Oct;9(5):436-447. doi: 10.1161/CIRCGENETICS.116.001506.
20. Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, Simonovic M, Doncheva NT, Morris JH, Bork P, Jensen LJ, Mering CV. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide

experimental datasets. Nucleic Acids Res. 2019 Jan 8;47(D1):D607-D613. doi: 10.1093/nar/gky1131.
21. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 2003 Nov;13(11):2498-504. doi: 10.1101/gr.1239303.

22. Knox C, Law V, Jewison T, Liu P, Ly S, Frolkis A, Pon A, Banco K, Mak C, Neveu V, Djoumbou Y, Eisner R, Guo AC, Wishart DS. DrugBank 3.0: a comprehensive resource for 'omics' research on drugs. Nucleic Acids Res. 2011 Jan;39(Database issue):D1035-41. doi: 10.1093/nar/gkq1126.