

## Identification and functional characterisation of DNA methylation differences between East- and West-originating Finns

Joanna Ciantar, Saara Marttila, Sonja Rajić, Daria Kostiniuk, Pashupati P Mishra, Leo-Pekka Lyytikäinen, Nina Mononen, Marcus E Kleber, Winfried März, Mika Kähönen, Olli Raitakari, Terho Lehtimäki & Emma Raitoharju

**To cite this article:** Joanna Ciantar, Saara Marttila, Sonja Rajić, Daria Kostiniuk, Pashupati P Mishra, Leo-Pekka Lyytikäinen, Nina Mononen, Marcus E Kleber, Winfried März, Mika Kähönen, Olli Raitakari, Terho Lehtimäki & Emma Raitoharju (2024) Identification and functional characterisation of DNA methylation differences between East- and West-originating Finns, *Epigenetics*, 19:1, 2397297, DOI: [10.1080/15592294.2024.2397297](https://doi.org/10.1080/15592294.2024.2397297)

**To link to this article:** <https://doi.org/10.1080/15592294.2024.2397297>



© 2024 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.



[View supplementary material](#)



Published online: 01 Sep 2024.



[Submit your article to this journal](#)



[View related articles](#)



[View Crossmark data](#)

## Identification and functional characterisation of DNA methylation differences between East- and West-originating Finns

Joanna Ciantar<sup>a</sup>, Saara Marttila<sup>a,b,c</sup>, Sonja Rajić<sup>a</sup>, Daria Kostiniuk<sup>a</sup>, Pashupati P Mishra<sup>d</sup>, Leo-Pekka Lyytikäinen<sup>d</sup>, Nina Mononen<sup>d</sup>, Marcus E Kleber<sup>e,f</sup>, Winfried März<sup>e,g</sup>, Mika Kähönen<sup>h</sup>, Olli Raitakari<sup>i,j,k</sup>, Terho Lehtimäki<sup>d,l</sup>, and Emma Raitoharju<sup>a,l,m</sup>

<sup>a</sup>Molecular Epidemiology (MOLE), Faculty of Medicine and Health Technology, Tampere University, Tampere, Finland; <sup>b</sup>Gerontology Research Center, Tampere University, Tampere, Finland; <sup>c</sup>Tays Research Services, Wellbeing Services County of Pirkanmaa, Tampere University Hospital, Tampere, Finland; <sup>d</sup>Department of Clinical Chemistry, Tays Research Services, Fimlab Laboratories, and Finnish Cardiovascular Research Center, Faculty of Medicine and Health Technology, Tampere University, Tampere, Finland; <sup>e</sup>Vth Department of Medicine (Nephrology, Hypertensiology, Endocrinology, Diabetology, Rheumatology), Medical Faculty of Mannheim, Heidelberg University, Mannheim, Germany; <sup>f</sup>SYNLAB MVZ Humangenetik Mannheim, Mannheim, Germany; <sup>g</sup>Synlab Academy, SYNLAB Holding Deutschland GmbH, Mannheim, Germany; <sup>h</sup>Department of Clinical Physiology, Tampere University Hospital and Faculty of Medicine and Health Technology, Tampere University, Tampere, Finland; <sup>i</sup>Centre for Population Health Research, University of Turku and Turku University Hospital, Turku, Finland; <sup>j</sup>Research Centre of Applied and Preventive Cardiovascular Medicine, University of Turku, Turku, Finland; <sup>k</sup>Department of Clinical Physiology and Nuclear Medicine, Turku University Hospital, Turku, Finland; <sup>l</sup>Finnish Cardiovascular Research Center Tampere, Faculty of Medicine and Health Technology, Tampere University, Tampere, Finland; <sup>m</sup>Fimlab Laboratories, Tampere, Finland

### ABSTRACT

Eastern and Western Finns show a striking difference in coronary heart disease-related mortality; genetics is a known contributor for this discrepancy. Here, we discuss the potential role of DNA methylation in mediating the discrepancy in cardiometabolic disease-risk phenotypes between the sub-populations. We used data from the Young Finns Study ( $n = 969$ ) to compare the genome-wide DNA methylation levels of East- and West-originating Finns. We identified 21 differentially methylated loci ( $FDR < 0.05$ ;  $\Delta\beta > 2.5\%$ ) and 7 regions (smoothed  $FDR < 0.05$ ;  $CpGs \geq 5$ ). Methylation at all loci and regions associates with genetic variants ( $p < 5 \times 10^{-8}$ ). Independently of genetics, methylation at 11 loci and 4 regions associates with transcript expression, including genes encoding zinc finger proteins. Similarly, methylation at 5 loci and 4 regions associates with cardiometabolic disease-risk phenotypes including triglycerides, glucose, cholesterol, as well as insulin treatment. This analysis was also performed in LURIC ( $n = 2371$ ), a German cardiovascular patient cohort, and results replicated for the association of methylation at cg26740318 and DMR\_11p15 with diabetes-related phenotypes and methylation at DMR\_22q13 with triglyceride levels. Our results indicate that DNA methylation differences between East and West Finns may have a functional role in mediating the cardiometabolic disease discrepancy between the sub-populations.

### ARTICLE HISTORY

Received 29 April 2024  
Revised 14 August 2024  
Accepted 22 August 2024

### KEYWORDS


DNA methylation; East and West Finns; population studies; multi-omic analysis; EWAS; quantitative trait loci

## Introduction

The Finnish population exhibits a striking division between the Northeast and the Southwest regions, characterised by differences in lifestyle, dialects, genetics and disease prevalence, including coronary heart disease (CHD) [1–4]. The East–West discrepancy in CHD mortality was first documented in the 1940s–50s and led to the launching of The North Karelian Project in 1972, with the aim of reducing CHD mortality in East Finland [5].

Previous research from The Young Finns Study (YFS), a longitudinal cohort set up with the aim of investigating these regional differences, identified higher total cholesterol, lower HDL-cholesterol and higher triglycerides in East-originating individuals in their childhood and adolescence [6,7]. Higher levels of Lp(a), consistent across all time points, were also identified in Eastern Finns [8]. Nationwide educational campaigns were successful in reducing CHD mortality across Finland [9,10];

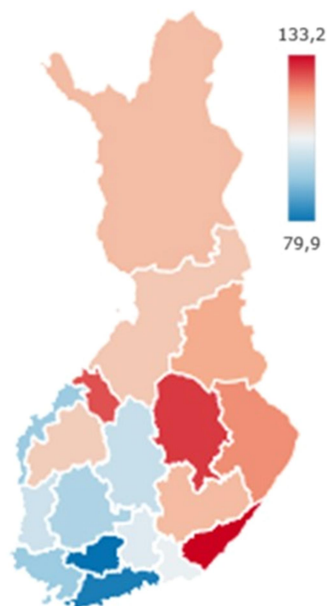
**CONTACT** Joanna Ciantar  [joanna.ciantar@tuni.fi](mailto:joanna.ciantar@tuni.fi); Emma Raitoharju  [emma.raitoharju@tuni.fi](mailto:emma.raitoharju@tuni.fi)  Molecular Epidemiology (MOLE), Faculty of Medicine and Health Technology, Tampere University, Ylpön katu 34, Tampere 33520, Finland

 Supplemental data for this article can be accessed online at <https://doi.org/10.1080/15592294.2024.2397297>

© 2024 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

## CHD index 2017 - 2019



**Figure 1.** Age-standardized coronary heart disease index in Finland. Data from the Finnish Institute for Health and Welfare shows the distribution of coronary heart disease (CHD) in the country. The index is calculated relative to the national average which is given a value of 100. Regions with an age-adjusted CHD index greater than the national average are redder in colour while those with a lower value are bluer, on a continuous colour scale. Map generated using the THL morbidity index online tool [15].

however, a relatively elevated level of CHD mortality in Northeast Finland remains to this day [11,12] (Figure 1). This residual risk difference could be partially attributed to differences in the genetic background of the two sub-populations [13,14].

Since previous research on cardiometabolic disease (CMD) risk differences between Eastern and Western Finns has centred around lifestyle and genetics, we wanted to investigate the potential role of DNA methylation in this context. In mammals, DNA methylation predominantly occurs on cytosine bases that are linked to an adjacent guanine base via a phosphate, therefore termed CpGs [16]. The DNA methylation profile changes with normal ageing processes [17,18] and in several diseases including cancer [19,20], neurological diseases [21,22] and CMDs [23–26]. DNA methylation differences between healthy individuals can also arise due to environmental or lifestyle differences, such as smoking [27–29] and obesity [30,31], as well as genetic variants, known as

methylation quantitative trait loci (meQTLs) [32,33]. Several studies have shown that genetic differences between populations significantly contribute to their different DNA methylation profiles [34,35]. The inheritance of meQTLs is a potential mechanism for the transmittance of cardiovascular disease risk across generations [36].

To investigate the potential role of DNA methylation in explaining the difference in CHD risk between Eastern and Western Finns, we utilised multi-omic data from the YFS cohort. We set out with three main aims: 1) to identify differentially methylated loci between East- and West-originating Finns 2) to identify meQTLs that might be regulating the differentially methylated loci and 3) to assess the role of DNA methylation at these loci in the regulation of gene expression and CMD-risk phenotypes. The Ludwigshafen Risk and Cardiovascular Health (LURIC) cohort was used as a replication cohort for the associations between methylation and CMD risk factors.

## Materials and methods

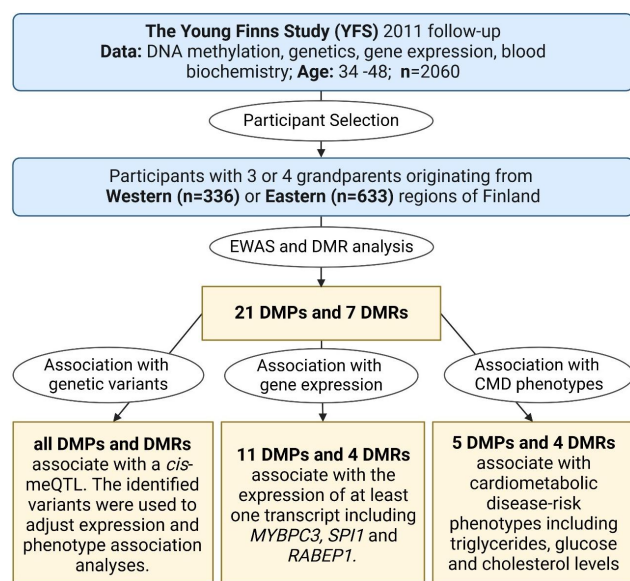
### Study population: young finns study (YFS)

YFS is a longitudinal multi-omic population cohort launched in 1980 with the aim of investigating regional differences in CHD risk factors in Finland [37]. The study participants ( $n = 3,596$ ), aged between 3 and 18 y, were recruited from five Finnish cities (Helsinki, Tampere, Turku, Oulu and Kuopio) and the nearby countryside. Study participants were sampled randomly from the Finnish national registry with an equal number of boys and girls and participants from Eastern and Western parts of the country. Follow-ups were performed every 3–9 y with the most recent one in 2018 which also included the parents and offspring of the original participants. Blood samples were taken by a uniform protocol at each of the five data collection sites by specially trained health professionals recruited for the study. The blood samples were frozen, transported to a centralised storage location and henceforth processed uniformly. Data utilized in this study was collected in the 2011 follow-up and includes DNA methylation, genotype, gene expression and biochemical blood measurements.

A subset of the YFS cohort, determined by the participant's self-reported grandparental birthplaces and the availability of DNA methylation data, was used to answer the primary research question presented in this study (Figure 2). Participants with at least three grandparents originating from Eastern or Western Finland were selected ( $n = 969$ ). A previously defined East–West demarcation [38] was used to categorize the Finnish regions.

### Replication cohort: the Ludwigshafen Risk and Cardiovascular health (LURIC)

LURIC is a German cardiovascular patient cohort recruited between 1997 and 2000 [39]. The participants had undergone coronary angiography at the Ludwigshafen hospital and were eligible for the study if they did not have any acute or chronic diseases other than cardiac diseases. Ages of the participants varied from 19 to 91



**Figure 2.** Flowchart of the study. Participants from the Young Finns Study (YFS) were selected to be included in this study based on their Eastern or Western origins, which were defined as having three or four grandparents originating from the same area. An epigenome-wide association study (EWAS) and Differentially Methylated Region (DMR) analysis revealed 21 differentially methylated positions (DMPs) and 7 DMRs between Eastern and Western originating Finns. Methylation at these loci was investigated for its association with underlying genetic variants, gene expression and cardiometabolic disease (CMD) risk phenotypes. The association between methylation at the identified loci and CMD risk phenotypes was replicated in LURIC, a German cardiovascular patient cohort (not shown in figure). Figure created with BioRender.com.

(mean ages of male and female participants were 65 and 62, respectively) at the time of recruitment. In the present work, this cohort is used as a replication cohort for the association between DNA methylation and cardiometabolic phenotypes.

### DNA methylation data

For YFS, DNA was extracted from EDTA-blood samples using the Wizard Genomic DNA Purification Kit, following the manufacturer's instructions. DNA methylation levels were measured using the Illumina Infinium MethylationEPIC BeadChip, following the Illumina protocol, for both cohorts. The array measures DNA methylation at approximately 850,000 CpG sites for a genome-wide coverage. The YFS samples from different collection sites were randomised on the arrays to avoid potential confounding batch effects.

DNA methylation data from YFS was pre-processed using the *minfi* R package [40] as previously described [41]. After preprocessing, 769,683 probes were retained, having excluded those which (i) target sites on sex chromosomes, (ii) are known to be cross-reactive or (iii) target known single nucleotide polymorphisms. Blood cell proportions were estimated with the Houseman method [42] using the *minfi* R package.

LURIC DNA methylation data was processed following the CPACOR pipeline [43] and has been described previously [44]. Briefly, probe exclusion was performed using the *DMRcate* R package [45] and 795,619 autosomal probes were retained. Methylation values were then quantile normalized. DNA methylation data from LURIC is available for 2371 participants.

### Genotyping

Genotyping of YFS participants was performed using a custom Illumina Human 670k BeadChip as described previously [46]. Genotype imputation was performed using TOPMed r2 as reference [47]. After QC checks, genotype data for 546,677 SNPs in 2443 individuals was retained. Genotyping in LURIC was performed with the Affymetrix Human SNP Array 6.0 [48]. After

QC checks, genotype data for 700,714 SNPs in 2947 individuals was retained. Imputation was performed with the 1000 Genomes Phase I integrated variant set [49] or TOPMed r2 as reference. This study utilized genetic data from both imputations as indicated in Table S1.

### Gene expression profiling

Whole blood for gene expression profiling was collected from YFS participants in PaXgene Blood RNA Tubes and RNA was isolated using the PAXgene Blood microRNA Kit. Genome-wide gene expression data was measured from blood samples collected in 2011 with the Illumina HumanHT-version 4 Expression BeadChip as described previously [50]. Raw expression data was exported from GenomeStudio and processed in R studio. Nonparametric background correction was performed, followed by quantile normalisation and  $\log_2$  transformation using the *limma* R package [51].

### Biochemical blood measurements

Phenotype data for 2063 YFS participants and 3316 LURIC participants was collected using a combination of physical examinations and questionnaires. Fasting blood samples were also collected at this time and used to measure glucose, insulin, triglycerides, lipoprotein cholesterol and C-reactive protein by standard clinical assays as previously described for YFS [52] and for LURIC [39].

### Identification of DMPs between Eastern and Western Finns

Classification of ‘Eastern’ and ‘Western’ Finnish regions was performed as described previously [38]. An epigenome-wide association study (EWAS) was performed using linear regression in R to determine differentially methylated CpG sites between East- and West-originating Finns (Figure 2). The model was adjusting with age, sex, BMI, smoking status, methylation-based estimation of blood cell composition and the first 30 principal components of the control probes. To account for multiple testing, p-values were adjusted with False Discovery Rate (FDR) correction. Difference in methylation ( $\Delta\beta$ ) between the groups was calculated for each CpG

site by subtracting the median value for the West-originating participants from the median value of the East-originating participants.

### Identification of DMRs between Eastern and Western Finns

Differentially methylated regions (DMRs) were identified using the *DMRcate* R package [45]. Beta methylation values were first converted to *M* values using the formula  $\log_2(\beta)/(1-\beta)$ . *M* methylation values were then used to calculate DMRs with  $\lambda = 1000$ ,  $C = 2$  and the same co-variables as described previously for EWAS. DMRs with a minimum of 5 CpG sites were retained for downstream analysis. The mean methylation level for each DMR was calculated for all participants, and this was used in downstream analyses of genetic, gene expression and phenotype association.

### Epigenetic clocks

Dunedin PACE epigenetic ages were calculated using the *DunedinPACE* R package [53], while principle component clocks and age acceleration residuals for Horvath, Hannum, Pheno Age and Grim Age [54] were calculated using the code provided by the authors. Inverse normal transformed epigenetic age residuals were compared between East- and West-originating Finns using regression models in R for all participants and separately for men and women. Regression models were adjusted with age, smoking and BMI.

### Genetic association studies

Genetic association studies were performed to identify *cis*-genetic variants (within 1Mb) that associate with DNA methylation (meQTL; Figure 2) and gene expression (eQTLs). All genetic association analyses were performed using PLINK 2.0 [55]. Statistically significant variants ( $p < 5e^{-8}$ ) with a minor allele frequency of  $\geq 0.02$  which passed the Hardy-Weinberg check ( $p \geq 1e^{-6}$ ) and had less than 5% missingness were identified. When multiple variants met these criteria, the variant with the lowest *p* value was selected as the top-associating variant.

### Gene expression association studies

Association analysis between methylation at previously identified DMPs and DMRs between East- and West-originating Finns and gene expression was performed on gene transcripts within 100kb of the methylation locus/region (Figure 2). Target genes were identified using the *biomaRt* R package [56]. Association analyses were performed using three linear regression models, the first containing no covariates, the second containing only the respective associating meQTL as a covariate and the third containing both the respective associating meQTL and eQTL.

### Association with cmd-related risk phenotypes

The analysis of the association between methylation at DMPs and DMRs and cardiometabolic phenotypes was performed with regression models in R (Figure 2). All models were adjusted with sex, age, smoking status and BMI for both YFS and LURIC, as well as fasting status upon blood collection for YFS. Inverse normal transformation was applied to all continuous dependent variables to normalize the variable distribution.

The continuous phenotypes included in the analyses for YFS were mean systolic and diastolic blood pressure (mmHg), glucose (mmol/l), insulin (mmol/l), HbA1c (mmol/mol), total cholesterol (mmol/l), LDL cholesterol (mmol/l), HDL cholesterol (mmol/l), triglycerides (mmol/l), apolipoprotein A-I (g/l), apolipoprotein B (g/l) and lipoprotein (a) (mg/l). Categorical phenotypes were hypertension, hypertension medication, hypercholesterolemia medication, insulin therapy, oral treatment for diabetes, diabetes type I and II, metabolic syndrome by harmonizing definition [57], cerebrovascular disease, heart bypass surgery, angioplasty on heart's blood vessels, heart failure, atrial fibrillation, other arrhythmia, congenital heart defect, cerebral thrombosis, cardiac infarction and coronary artery-induced chest pain.

The phenotypes included in the analyses for LURIC were mean systolic blood pressure (mmHg), mean diastolic blood pressure (mmHg), fasting blood glucose (mg/dL), fasting insulin (U/l), triglycerides (mg/dL), total cholesterol (mg/dL), LDL cholesterol (mg/dL), HDL cholesterol

(mg/dL), HbA1c (%), lipoprotein (a) (mg/dL), Apolipoprotein B (mg/dL), myocardial infarction, history of hypertension, diabetes type 2, insulin treatment and venous thrombosis.

## Results

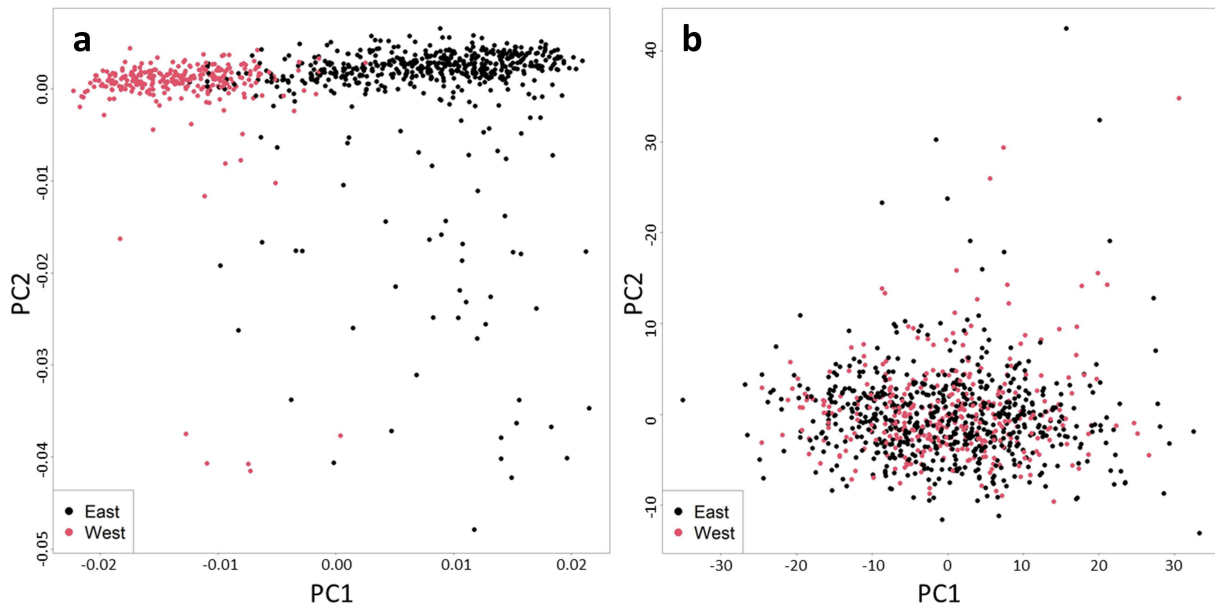
### The study population

In this study, we used a subset of the YFS ( $n = 969$ ) which included Finnish participants with Eastern or Western origins based on self-reported grandparental birthplace (Figure 2). Around half of the participants with Eastern origins were born and were also living in Eastern regions in 2011, while the rest migrated to the West. In contrast, almost all of the participants with Western origins were born and were living in the West in 2011 (Fig S1). The two groups show comparable demographics (Table S2).

In order to validate our classification of East/West origins, we analysed the first two principal components of our genetic data (available for 822 out of the 969 participants included in this analysis). As expected, the greatest source of variation in the genetic data (PC1) is explained by the East/West origin of the participants (Mann-Whitney U-test  $p$ -value  $< 2.2 \times 10^{-16}$ ), evident by the clustering of the two groups (Figure 3a). A similar analysis of principal components of DNA methylation data, however, did not show similar clustering (Figure 3b), indicating that the East/West origins of participants was not among the top causes of DNA methylation variation in 2011 (Mann-Whitney U-test  $p$ -value for PC1 = 0.838 and PC2 = 0.13).

### DNA methylation differences between East- and West-originating Finns

To identify DMPs between East- and West-originating Finns, we performed an epigenome-wide association study (EWAS). The analysis revealed 82 statistically significant DMPs at FDR  $< 0.05$  (Table S3). Of these, 21 CpGs had a difference in median methylation  $\beta$  value of  $\geq 2.5\%$  between the two groups (Table 1), which increases the potential biological relevance of these loci. These CpG sites were therefore selected for further analysis. Eastern Finns had



**Figure 3.** Variation in genetic and DNA methylation data of study participants. Scatter plots of the first two principal components of the genetic (A) and DNA methylation (B) data of the study cohort reveal a distinct clustering of East- and West-originating Finns only in the genetic data.

higher methylation levels at 14 of these CpG sites, while Western Finns had higher levels in 7 CpG sites. Most CpG sites are located in inter-genic regions, while nine sites are located in gene bodies and four are in transcription start sites. Several DMPs identified between Eastern and Western Finns exhibit notable clustering, particularly in chromosomes 2 and 17.

In addition to DMPs, we also wanted to identify larger differentially methylated regions (DMRs) as these may be more biologically relevant than individual CpG sites [58]. We identified seven DMRs ( $\geq 5$  CpGs per region) between East- and West-originating Finns (Table 2). The regions span from 325 to 924 basepairs in length and contain up to 21 CpG sites. For the scope of

**Table 1.** The 21 DMPs between East- and West-originating Finns.

CpG	Chr	Locus	Gene Name	Relation to Gene	Beta	SE	P value	FDR	$\Delta\beta$
cg08946995	1	229283674	/	/	-0.03	0.006	$1.3 \times 10^{-6}$	$3.1 \times 10^{-2}$	5.1%
cg15720821	1	229284563	/	/	-0.03	0.007	$8.7 \times 10^{-7}$	$2.7 \times 10^{-2}$	3.5%
cg13902335	2	216876969	<i>MREG</i>	Body	0.02	0.004	$1.5 \times 10^{-7}$	$1.4 \times 10^{-2}$	-3.1%
cg10833299	2	242749776	<i>NEU4</i>	TSS1500	-0.05	0.008	$4.9 \times 10^{-9}$	$9.3 \times 10^{-4}$	8.1%
cg15470645	2	242749789	<i>NEU4</i>	TSS1500	-0.04	0.006	$1.0 \times 10^{-10}$	$7.9 \times 10^{-5}$	5.5%
cg26207766	2	242763794	/	/	-0.06	0.011	$1.0 \times 10^{-7}$	$1.3 \times 10^{-2}$	5.8%
cg03455424	2	242763982	/	/	-0.04	0.007	$1.6 \times 10^{-9}$	$5.6 \times 10^{-4}$	7.3%
cg18948877	2	242904738	<i>LINC01237</i>	Body	-0.02	0.004	$3.0 \times 10^{-7}$	$1.9 \times 10^{-2}$	2.5%
cg03514530	3	145525915	/	/	0.04	0.009	$3.5 \times 10^{-6}$	$4.4 \times 10^{-2}$	-6.4%
cg10619644	7	69149951	<i>AUTS2</i>	Body	0.03	0.006	$1.9 \times 10^{-6}$	$3.2 \times 10^{-2}$	-3.3%
cg26740318	11	18477153	<i>LDHAL6A</i>	TSS1500	-0.03	0.006	$4.0 \times 10^{-6}$	$4.5 \times 10^{-2}$	3.6%
cg00674440	11	47339067	<i>MADD</i>	Body	0.05	0.009	$5.8 \times 10^{-7}$	$2.5 \times 10^{-2}$	-8.6%
cg11472901	13	110527539	/	/	0.02	0.005	$2.9 \times 10^{-6}$	$3.8 \times 10^{-2}$	-3.0%
cg27489373	16	1507959	<i>CLCN7</i>	Body	-0.03	0.007	$5.3 \times 10^{-6}$	$5.0 \times 10^{-2}$	2.8%
cg23194629	17	5290760	<i>NUP88</i>	Body	-0.02	0.005	$3.6 \times 10^{-6}$	$4.4 \times 10^{-2}$	3.1%
cg01199135	17	71471080	<i>SDK2</i>	Body	-0.02	0.004	$7.5 \times 10^{-8}$	$1.2 \times 10^{-2}$	3.0%
cg09786634	17	78912393	<i>RPTOR</i>	Body	0.03	0.006	$5.8 \times 10^{-7}$	$2.5 \times 10^{-2}$	-3.0%
cg11949518	17	78912765	<i>RPTOR</i>	Body	0.06	0.012	$4.2 \times 10^{-6}$	$4.7 \times 10^{-2}$	-3.1%
cg00698263	17	78962984	/	/	-0.02	0.005	$4.5 \times 10^{-6}$	$4.8 \times 10^{-2}$	3.9%
cg03953789	17	78963035	/	/	-0.02	0.005	$4.6 \times 10^{-6}$	$4.8 \times 10^{-2}$	6.6%
cg14750269	19	46975180	<i>PNMAL1</i>	TSS1500	-0.02	0.004	$8.8 \times 10^{-7}$	$2.7 \times 10^{-2}$	3.6%

Significant DMPs (FDR  $< 0.05$  and  $\Delta\beta \geq 2.5\%$ ) between East- and West-originating Finns.  $\Delta\beta$  is the difference in median methylation  $\beta$  value between the two groups ( $\beta_{\text{East}} - \beta_{\text{West}}$ ). The Beta column represents the regression coefficient. Body – CpG site located within the gene body; TSS1500 – CpG site located within 1500bp of the transcription start site; Shore – CpG site is located up to 2 kb from CpG island; Shelf – CpG site is located between 2 and 4 kb from CpG island. The table is sorted by genomic locations (GRCh37 build).

this article, the DMRs were named according to the chromosome band in which they are located. DMR\_2q37 overlaps with two DMPs (cg10833299 and cg15470645), while the DMR\_11p15 overlaps with one DMP (cg26740318) that were identified in the EWAS. Eastern Finns are hypermethylated for all DMRs when compared to Western Finns (Fig S2), a trend which follows the EWAS results.

### Global methylation and epigenetic age analysis

Next, we investigated whether there are differences in global methylation and epigenetic ages between East- and West-originating Finns. There was no statistically significant difference in global methylation between the two groups ( $p = 0.12$ ; Fig S3). Epigenetic ages are an indicator of the biological health status and were calculated for DunedinPACE [53], a measure of the pace of aging, and PC clocks for Horvath, Hannum, PhenoAge and GrimAge [54]. Whole population and sex-stratified analysis revealed a statistically significant difference only in DunedinPACE ages of men, where East-originating men had a median value which was 0.02 higher than West-originating men ( $p = 0.014$ ; Table S4).

### The genetic basis of DMPs and DMRs between East- and West-originating Finns

Since DNA methylation levels can be regulated both by genetic and environmental factors, we investigated whether methylation at each of the 21 identified DMPs associates with genetic variants within 1Mb of the locus. Multiple *cis*-meQTLs were identified for each CpG site ( $p < 5e^{-8}$ ; Table S5) and the variant with the lowest p-value was selected for further analysis. Similarly, mean

methylation level of each DMR is associated with multiple genetic variants.

The strong associations of DMP and DMR methylation levels with genetics raised the question whether the observed DNA methylation differences between Eastern and Western Finns are due to an unequal distribution of genetic variants. To investigate this, we replicated the initial EWAS setting for the 21 DMPs with the addition of the respective associating top SNP as a covariate. The top SNP was considered to account for the genetic effect since there is high association between proximal SNPs due to linkage disequilibrium. Our results show a decrease in overall statistical significance when compared to the original EWAS setting (Table 3). However, the methylation at 13 CpG sites remained nominally significant after adjusting with the respective SNP, indicating that differential DNA methylation at these sites is not purely dependent on the underlying genetics, and other factors – such as environmental factors – could be involved. Interestingly, the clustering CpG sites on chromosome 2 all retained statistical significance while those on chromosome 17 did not.

### Association of methylation at DMPs and DMRs with proximal gene expression

Next, we investigated whether DNA methylation at the DMPs associates with the expression of transcripts located 100kb upstream and downstream of the methylation locus. We found significant ( $p < 0.001$ ) associations between 13 DMPs and transcripts from 17 genes (Table S6). To account for the genetic effect on DNA methylation, we repeated the analysis and added the respective meQTL as a covariate. The association between 10 CpG sites and transcripts from 12 genes retained nominal statistical significance (Table 4A). We also made a third model which

**Table 2.** Differentially methylated regions between Eastern and Western Finns.

Chr band	Start Position	End Position	Width	No. Of CpGs	Overlapping Genes	Smoothed FDR	Mean $\Delta\beta$	Max $\Delta\beta$
2q37	242749533	242749925	393	5	<i>NEU4, AC114730.3</i>	$1.4 \times 10^{-14}$	2.9%	8.2%
6p22	29648161	29649084	924	21	<i>ZFP57</i>	$5.3 \times 10^{-23}$	1.7%	5.1%
6p21	32121113	32121566	454	19	<i>PPT2, PRRT1</i>	$8.5 \times 10^{-17}$	0.5%	1.2%
8p21	22132563	22133356	794	13	<i>PIWIL2</i>	$8.5 \times 10^{-17}$	0.7%	2.4%
11p15	18477153	18477534	382	5	<i>LDHAL6A</i>	$3.1 \times 10^{-14}$	1.7%	3.6%
19q13	58220494	58220818	325	8	<i>ZNF551, AC003006.7, ZNF154</i>	$8.9 \times 10^{-14}$	0.8%	1.7%
22q13	38092719	38093207	489	10	<i>NOL12, TRIOBP</i>	$5.3 \times 10^{-14}$	0.7%	1.2%

DMR analysis revealed seven DMRs, containing  $\geq 5$  CpGs, between East- and West-originating Finns.  $\Delta\beta$  is the difference in median methylation  $\beta$  value between the two groups ( $\beta_{\text{East}} - \beta_{\text{West}}$ ). Start and End positions for the DMR are based on the GRCh37 build.

accounts for the effect of eQTLs, genetic variants that associate with gene expression, in addition to meQTL. A nominally significant association was retained for all previously identified associations except that of cg11949518 and *RPTOR* expression, and cg01199135 and *SDK2* expression. These results suggest that DNA methylation at the sites which retained statistical significance is likely to be mechanistically involved in gene expression regulation and not a bystander.

As with the DMPs, we investigated whether methylation at the DMRs associates with proximal gene expression. Results show that the methylation at six out of the seven DMRs associates with at least one transcript without adjusting with genetics (Table S6). After adjusting with the respective meQTL, the association between 4 DMRs and 15 transcripts retained nominal statistical significance (Table 4B). Further adjusting the model with both meQTL and eQTL did not drastically affect the statistical significance of the results.

### Association of DNA methylation at DMPs and DMRs with cardiometabolic phenotypes

Next, we assessed whether the differential methylation between Eastern and Western Finns could potentially affect cardiometabolic health. Out of the 21 DMPs tested, we found associations with

CMD risk phenotypes, independent of genetics, for 5 DMPs and 4 DMRs (Table 5). Associating CMD-risk phenotypes include triglycerides, glucose, total cholesterol and LDL cholesterol, as well as insulin treatment and a metabolic syndrome diagnosis following harmonized criteria [57]. The full results of nominally significant associations for DMPs and DMRs before and after adjusting with genetics can be found in Table S7.

Since the YFS cohort is a population cohort consisting of relatively young and healthy individuals (34–48 y old in 2011), most participants had not yet started showing severe CMD phenotypes and so, effects of altered DNA methylation patterns may not be seen. We therefore replicated the analysis of association between methylation and CMD risk phenotype in LURIC, a cardiovascular patient cohort from Germany. Out of the 21 DMPs and 7 DMRs tested, 8 DMPs and 3 DMRs associated with one or more CMD-risk phenotype independently of genetics (Table 6). Full results before and after adjusting with genetics can be found in Table S8. Numerous loci associated with triglyceride, HbA1c and cholesterol levels as well as fasting insulin ( $p < 0.05$ ). Methylation at both cg26740318 and DMR\_11p15, in which the CpG site lies, associates with fasting insulin levels. Interestingly, cg11472901, cg18948877 (located in the *LINC01237* gene body) and cg26740318

**Table 3.** Adjusting the EWAS analysis with the top-associating *cis*-meQTL.

CpG	Chr	Locus	Gene Name	Top SNP	Beta	SE	P value
cg08946995	1	229283674	/	rs11583419	<0.001	0.002	$8.6 \times 10^{-1}$
cg15720821	1	229284563	/	rs72766103	-0.008	0.004	$8.6 \times 10^{-2}$
cg13902335	2	216876969	<i>MREG</i>	rs13013441	0.007	0.003	<b><math>2.4 \times 10^{-2}</math></b>
cg10833299	2	242749776	<i>NEU4</i>	rs4973673	-0.032	0.008	<b><math>8.4 \times 10^{-5}</math></b>
cg15470645	2	242749789	<i>NEU4</i>	rs34334574	-0.029	0.006	<b><math>1.1 \times 10^{-5}</math></b>
cg26207766	2	242763794	/	rs34334574	-0.038	0.012	<b><math>1.0 \times 10^{-3}</math></b>
cg03455424	2	242763982	/	rs34334574	-0.030	0.007	<b><math>5.7 \times 10^{-5}</math></b>
cg18948877	2	242904738	<i>LINC01237</i>	rs4973673	-0.009	0.004	<b><math>2.4 \times 10^{-2}</math></b>
cg03514530	3	145525915	/	rs7643638	0.018	0.006	<b><math>2.8 \times 10^{-3}</math></b>
cg10619644	7	69149951	<i>AUTS2</i>	rs56302473	0.008	0.003	<b><math>1.7 \times 10^{-2}</math></b>
cg26740318	11	18477153	<i>LDHAL6A</i>	rs34109624	-0.014	0.007	<b><math>3.5 \times 10^{-2}</math></b>
cg00674440	11	47339067	<i>MADD</i>	rs7947269	0.008	0.003	<b><math>2.2 \times 10^{-2}</math></b>
cg11472901	13	110527539	/	rs12583851	0.016	0.005	<b><math>3.9 \times 10^{-3}</math></b>
cg27489373	16	1507959	<i>CLCN7</i>	rs9941109	-0.022	0.007	<b><math>2.3 \times 10^{-3}</math></b>
cg23194629	17	5290760	<i>NUP88</i>	rs60379759	0.002	0.002	$3.1 \times 10^{-1}$
cg01199135	17	71471080	<i>SDK2</i>	rs11651551	-0.016	0.004	<b><math>2.0 \times 10^{-4}</math></b>
cg09786634	17	78912393	<i>RPTOR</i>	rs7219318	0.004	0.003	$2.5 \times 10^{-1}$
cg11949518	17	78912765	<i>RPTOR</i>	rs56239682	0.010	0.006	$7.8 \times 10^{-2}$
cg00698263	17	78962984	/	rs6565507	-0.003	0.003	$3.8 \times 10^{-1}$
cg03953789	17	78963035	/	rs1555643536	-0.002	0.003	$5.6 \times 10^{-1}$
cg14750269	19	46975180	<i>PNMAL1</i>	rs11083832	-0.003	0.002	$2.3 \times 10^{-1}$

A recreation of the original EWAS setting with the addition of the respective top SNP as a covariate resulted in 13 CpGs which retained nominal statistical significance (in bold). The chromosome, locus (GRCh37 build) and gene name pertain to the CpG site while the Top SNP column lists the rsID of the SNP added as a covariate.

**Table 4.** Association of methylation at DMPs and DMRs with proximal gene expression.

A	CpG	Gene	Model 1: No covariates		Model 2: meQTL		Model 3: meQTL + eQTL	
			Beta	P value	Beta	P value	Beta	P value
			cg00674440	<i>DDB2</i>	0.87	$3.7 \times 10^{-5}$	2.62	$4.9 \times 10^{-6}$
	<i>MYBPC3</i>	-2.27	$7.0 \times 10^{-28}$	-6.83	$2.1 \times 10^{-35}$	-6.16	$5.6 \times 10^{-31}$	
	<i>SPI1</i>	-1.72	$2.0 \times 10^{-16}$	-5.21	$1.7 \times 10^{-20}$	-4.32	$2.3 \times 10^{-16}$	
		-0.73	$5.8 \times 10^{-4}$	-5.20	$3.8 \times 10^{-20}$	-5.43	$3.8 \times 10^{-24}$	
cg01199135	<i>SDK2</i>	-1.98	$1.9 \times 10^{-5}$	-2.82	$7.2 \times 10^{-7}$	-0.63	$2.8 \times 10^{-1}$	
cg03455424	<i>FLJ33590</i>	1.14	$1.2 \times 10^{-5}$	0.97	$3.8 \times 10^{-3}$	/	/	
cg09786634	<i>RPTOR</i>	1.97	$5.4 \times 10^{-9}$	2.82	$1.4 \times 10^{-5}$	2.71	$2.4 \times 10^{-5}$	
cg10833299	<i>FLJ33590</i>	0.89	$1.4 \times 10^{-4}$	0.66	$3.7 \times 10^{-2}$	/	/	
cg11949518	<i>CHMP6</i>	1.05	$1.1 \times 10^{-10}$	1.69	$2.4 \times 10^{-5}$	0.87	$2.8 \times 10^{-2}$	
	<i>FLJ90757</i>	1.16	$8.4 \times 10^{-13}$	1.10	$5.6 \times 10^{-3}$	0.80	$3.9 \times 10^{-2}$	
	<i>RPTOR</i>	0.81	$6.4 \times 10^{-7}$	1.21	$2.6 \times 10^{-3}$	0.70	$8.4 \times 10^{-2}$	
cg14750269	<i>PNMAL1</i>	-3.61	$7.9 \times 10^{-15}$	-4.78	$1.2 \times 10^{-5}$	-3.28	$4.0 \times 10^{-3}$	
cg18948877	<i>FLJ33590</i>	-2.25	$3.8 \times 10^{-6}$	-3.45	$1.1 \times 10^{-9}$	/	/	
cg23194629	<i>DHX33</i>	1.31	$1.5 \times 10^{-4}$	2.79	$1.6 \times 10^{-5}$	/	/	
	<i>RABEP1</i>	6.29	$1.6 \times 10^{-83}$	3.66	$3.5 \times 10^{-12}$	3.65	$2.2 \times 10^{-12}$	
	<i>RPAIN</i>	-1.20	$5.5 \times 10^{-4}$	9.01	$1.0 \times 10^{-47}$	/	/	
		1.35	$9.4 \times 10^{-5}$	3.45	$8.3 \times 10^{-8}$	/	/	
cg26207766	<i>FLJ33590</i>	0.82	$4.7 \times 10^{-6}$	0.67	$2.3 \times 10^{-3}$	/	/	
B	DMR	Gene	Beta	P value	Beta	P value	Beta	P value
	2q37	<i>FLJ33590</i>	2.63	$1.0 \times 10^{-4}$	1.97	$2.7 \times 10^{-2}$	/	/
	19q13	<i>ZNF134</i>	-3.17	$2.3 \times 10^{-7}$	-3.15	$1.9 \times 10^{-6}$	/	/
		<i>ZNF154</i>	-8.87	$5.7 \times 10^{-51}$	-8.50	$1.9 \times 10^{-40}$	-7.96	$4.1 \times 10^{-37}$
		<i>ZNF211</i>	-3.74	$9.1 \times 10^{-10}$	-4.03	$9.3 \times 10^{-10}$	-4.21	$6.2 \times 10^{-11}$
		<i>ZNF586</i>	-2.23	$2.8 \times 10^{-4}$	-2.38	$3.3 \times 10^{-4}$	/	/
		<i>ZNF671</i>	-6.18	$1.4 \times 10^{-24}$	-6.06	$9.9 \times 10^{-21}$	-5.65	$1.8 \times 10^{-18}$
		<i>ZNF776</i>	-3.65	$2.4 \times 10^{-9}$	-3.36	$3.7 \times 10^{-7}$	/	/
	22q13	<i>LGALS1</i>	2.64	$5.3 \times 10^{-4}$	3.29	$1.9 \times 10^{-4}$	3.61	$3.6 \times 10^{-5}$
		<i>NOL12</i>	-2.85	$1.8 \times 10^{-4}$	-2.28	$9.8 \times 10^{-3}$	/	/
		<i>PDXP</i>	2.95	$1.1 \times 10^{-4}$	2.79	$1.6 \times 10^{-3}$	/	/
		<i>TRIOBP</i>	-5.97	$2.8 \times 10^{-15}$	-3.51	$3.2 \times 10^{-5}$	-2.80	$6.8 \times 10^{-4}$
	6p21	<i>NOTCH4</i>	-3.83	$1.2 \times 10^{-4}$	-2.85	$2.5 \times 10^{-2}$	/	/

Three regression models were used to identify associations between DNA methylation at DMPs (A) and DMRs (B) with proximal gene expression. The first model contained no covariates, while the second model contained only the respective meQTL (methylation quantitative trait locus) as a covariate and the third model included both meQTL and eQTL (expression QTL). Only the results for associations that were significant ( $p < 0.001$ ) in the first model and nominally significant in the second model are shown (see Table S6 for full results). Missing values for the third column are due to lack of an identifiable eQTL.

(located in the *LDHAL6A* TSS1500 region) as well as DMR\_11p15, DMR\_19q13 and DMR\_22q13 associate with phenotypes in both YFS and LURIC. Of these, associating phenotypes between cohorts is remarkably similar for cg26740318 and DMR\_11p15 which associate with insulin treatment in YFS and fasting insulin levels in LURIC and DMR\_22q13 which associates with triglyceride levels in both cohorts.

## Discussion

We have shown for the first time that East- and West-originating Finns exhibit distinct DNA methylation profiles. We identified 21 DMPs (FDR <0.05 and  $\Delta\beta > 2.5\%$ ), as well as 7 DMRs between Finns

originating from Eastern or Western regions. Furthermore, we have exploited the multiple omics available in the YFS cohort to gain insight to the establishment and potential function of the methylation at these DMPs and DMRs. All the identified DMPs and DMRs have a strong genetic basis; however, methylation at 13 DMPs may also be partly influenced by other factors. Independently of genetics, methylation at 11 DMPs and 4 DMRs associates with gene expression and 5 DMPs and 4 DMRs with cardiometabolic phenotypes in YFS. Similarly, methylation at 8 DMPs and 3 DMRs associates with cardiometabolic phenotypes independently of genetics in LURIC, a German cardiovascular patient cohort, with the association replicating for cg26740318, DMR\_11p15 and DMR\_22q13 across cohorts.

**Table 5.** Association of methylation at DMPs and DMRs with cmd-risk phenotypes.

A	CpG	Gene Name	Phenotype	Beta	CI2.5	CI97.5	SE	P value
cg03455424	/	/	LDL cholesterol	0.82	0.15	1.48	0.34	$1.6 \times 10^{-2}$
			Total cholesterol	0.80	0.14	1.45	0.33	$1.7 \times 10^{-2}$
cg11472901	/	/	Glucose	-1.29	-2.14	-0.45	0.43	$2.8 \times 10^{-3}$
			Hypercholesterolemia medication	-4.85	-9.51	-0.23	2.36	$4.0 \times 10^{-2}$
			Metabolic syndrome	-3.96	-6.82	-1.13	1.45	$6.3 \times 10^{-3}$
			Triglycerides	-1.24	-2.04	-0.43	0.41	$2.6 \times 10^{-3}$
cg18948877	LINC01237	Glucose	-1.30	-2.35	-0.25	0.53	$1.5 \times 10^{-2}$	
cg26740318	LDHAL6A	Insulin treatment	13.84	1.91	27.40	6.45	$3.2 \times 10^{-2}$	
cg27489373	CLCN7	/	Apolipoprotein B	0.88	0.29	1.47	0.30	$3.6 \times 10^{-3}$
			Triglycerides	1.07	0.49	1.64	0.29	$2.9 \times 10^{-4}$
B	DMR	Gene Name	Phenotype	Beta	CI2.5	CI97.5	SE	P value
11p15	/	LDHAL6A	Insulin treatment	29.54	6.51	56.72	12.79	$2.1 \times 10^{-2}$
			Systolic blood pressure	-1.06	-2.07	-0.06	0.51	$3.8 \times 10^{-2}$
19q13	ZNF551, AC003006.7, ZNF154	Metabolic syndrome	-5.61	-9.77	-1.55	2.09	$7.4 \times 10^{-3}$	
22q13	NOL12, TRIOBP	Triglycerides	1.69	0.20	3.17	0.76	$2.6 \times 10^{-2}$	
6p21	PPT2, PRRT1	Lipoprotein(a)	2.57	0.07	5.07	1.28	$4.4 \times 10^{-2}$	

The associations between methylation at DMPs (A) and DMRs (B) and CMD-related phenotype were adjusted with sex, age, BMI, smoking, fasting status and meQTL (methylation Quantitative Trait Locus). Only the nominally significant results after adjusting with meQTL are shown in this table, see Table S7 for full results. LDL cholesterol, Total cholesterol, Glucose and Triglycerides were measured in mmol/l. Apolipoprotein B was measured in g/l. Systolic blood pressure was measured in mmHg. Lipoprotein (a) was measured in mg/l.

In our study setting, we used the participant's grandparental birthplace to define their Eastern or Western origins. Since internal migration in Finland was very limited before the mid-1900s [59], the grandparental birthplace of our participants (aged 30–46 in 2011) is indicative of a long family history in the same region. Limited internal migration over a long period of time most likely resulted in genetic drift which led to the observed genetic discrepancy between the sub-populations [3,60–63]. These findings are also reflected in the YFS cohort genetic data, whose greatest source of

variation can be explained by the East/West origins of the participants.

Among the most notable environmental differences between North-Eastern and South-Western Finnish regions are differences in socio-economic status and urbanization [64–66]. Both socio-economic status and urbanization have been associated with cardiometabolic health outcomes [64,67,68] and DNA methylation has been proposed as the molecular mediator for this environmental effect on health [69–71]. For example, lower education

**Table 6.** Association of methylation at DMPs and DMRs with cmd-risk phenotypes in LURIC.

A	CpG	Gene Name	Phenotype	Beta	CI2.5	CI97.5	SE	P value
cg00674440	/	MADD	Triglycerides	0.92	0.46	1.39	0.24	$1.2 \times 10^{-4}$
cg10619644	/	AUTS2	HbA1c	0.91	0.25	1.56	0.33	$6.6 \times 10^{-3}$
cg11472901	/	/	Venous thrombosis	2.39	0.13	4.67	1.16	$3.9 \times 10^{-2}$
cg13902335	/	MREG	Fasting insulin	-0.78	-1.53	-0.03	0.38	$4.2 \times 10^{-2}$
			HbA1c	-1.18	-1.95	-0.40	0.40	$3.1 \times 10^{-3}$
cg15720821	/	/	Triglycerides	0.71	0.13	1.28	0.29	$1.6 \times 10^{-2}$
cg18948877	LINC01237	/	Diastolic blood pressure	-1.39	-2.48	-0.31	0.55	$1.2 \times 10^{-2}$
			Total cholesterol	-1.38	-2.47	-0.29	0.55	$1.3 \times 10^{-2}$
			Triglycerides	-1.49	-2.56	-0.43	0.54	$6.1 \times 10^{-3}$
cg23194629	NUP88	/	Total cholesterol	0.58	0.02	1.14	0.29	$4.4 \times 10^{-2}$
			Triglycerides	1.30	0.76	1.85	0.28	$3.1 \times 10^{-6}$
cg26740318	LDHAL6A	/	Fasting insulin	-0.51	-0.86	-0.17	0.18	$3.6 \times 10^{-3}$
			Triglycerides	-0.59	-0.95	-0.23	0.18	$1.4 \times 10^{-3}$
B	DMR	Gene Name	Phenotype	Beta	CI2.5	CI97.5	SE	P Value
11p15	/	LDHAL6A	Fasting insulin	-0.53	-1.03	-0.02	0.26	$4.0 \times 10^{-2}$
			LDL cholesterol	1.11	0.30	1.92	0.41	$7.3 \times 10^{-3}$
19q13	ZNF551, AC003006.7, ZNF154	Triglycerides	1.47	0.56	2.38	0.47	$1.6 \times 10^{-3}$	
22q13	NOL12, TRIOBP	LDL cholesterol	-1.25	-2.18	-0.32	0.47	$8.5 \times 10^{-3}$	

The associations between methylation at DMPs (A) and DMRs (B) and CMD-related phenotype in LURIC, a German cardiovascular patient cohort. Regression models were adjusted with sex, age, BMI, smoking and meQTL (methylation Quantitative Trait Locus). Only the nominally significant results after adjusting with meQTL are shown in this table, see Table S8 for full results. LDL cholesterol, Total cholesterol and Triglycerides were measured in mg/dL. Diastolic blood pressure was measured in mmHg Fasting insulin measurements are in U/l. HbA1c measurements are in %.

attainment is associated with both increased cardiovascular disease [72] and increased epigenetic age [71], although the exact role of the epigenetic changes remains unclear. The difference in socio-economic status and urbanization could be potential causes for the genetic independent observed differences in DNA methylation between East- and West-originating identified in this study.

The identified DMPs are located in or near 10 protein-coding genes and in one long non-coding RNA (*LINC01237*) with notable clustering in chromosomes 2 (6 DMPs) and 17 (6 DMPs). The locus with the highest statistical difference, in addition to more than 5% difference in the methylation  $\beta$ , between East- and West-originating Finns was cg15470645, located in the TSS1500 of *NEU4*, where another identified DMP (cg10833299) is also located. Both these CpGs are also within DMR\_2q37, which was found to be hypermethylated in the population originating from Eastern Finland. *NEU4* encodes a sialidase which has been implicated in cancer, lysosomal storage disease, neurogenesis and apoptosis [73] and intriguingly, methylation at this region has been associated with prenatal events [74,75].

Similarly, the methylation at two DMPs (cg09786634 and cg11949518) in the *RPTOR* gene body was significantly down-regulated in Eastern Finns compared to Western Finns. This is particularly interesting as *RPTOR* encodes for raptor (regulatory associated protein of mTOR), which is an integral part of the mTOR complex 1 [76,77], known to be involved in numerous important cellular processes including cell growth, metabolism, lysosome biogenesis and autophagy [78]. Previous research has identified altered DNA methylation patterns in *RPTOR* in lung cancer [79] and breast cancer [80,81] patients, although some of these results failed to replicate [82]. Mouse models have also shown that mTOR1 inhibition increases lifespan and improves cardiac function in a disease background [83]. A previous study identified three other CpGs in the same region which are hypermethylated in atherosclerotic aorta samples [84].

Out of the seven DMRs identified in this study, DMR\_6p22 and DMR\_19q13 overlap with genes encoding Krüppel-associated box (KRAB) domain containing zinc finger proteins. DMR\_6p22

overlaps with *ZFP57*, which is important for the maintenance of methylation in imprinting control regions [85–87]. A previous study found that maternal folate intake during pregnancy associates with the methylation of a region that largely overlaps the DMR\_6p22 region identified in this study [88]. On the other hand, DMR\_19q13 overlaps with *ZNF551* and *ZNF154*, and DNA methylation in three CpGs in *ZNF154* previously showed a positive correlation with age [89]. In our data, methylation at DMR\_19q13 also associates strongly with the expression of multiple zinc finger proteins, which have many diverse functions including transcriptional regulation, post-transcriptional modifications and protein–protein interactions [90]. Several zinc finger proteins play a role in cardiometabolic health and have been implicated in lipid metabolism [91,92], diabetes [93–95] and congenital heart diseases [96–98].

In addition to differences in DMPs and DMRs, a 0.02 increase in epigenetic age was identified in Eastern men when compared to their Western counterparts using the DunedinPACE metric but not with any of the other epigenetic age clocks included in this analysis. The DunedinPACE clock is the only clock trained on a longitudinal data and it yields a measure of the pace of ageing [53] rather than the biological age, making it unique from the other epigenetic age measures. A value of 1 in DunedinPACE indicates biological ageing of 1 y within a calendar year and thus, a higher pace of ageing is associated with a faster decline of physical and cognitive health [99,100]. Our finding is in line with other literature suggesting that Eastern Finns are worse off health-wise than Western Finns [5,11,12].

The mechanistic origin of DNA methylation differences between Eastern and Western Finns is made complicated due to the dual effects of genetic and environmental factors on the establishment of DNA methylation levels. Differential methylation between the two populations could be caused by different distribution of meQTLs across the sub-populations, as was previously reported in studies among genetically diverse populations [34,35]. The proposed mechanism of *cis*-meQTL action is through the disruption of CTCF-binding factor or transcription factor binding which then affects nearby DNA methylation [101]. Multiple *cis*-meQTLs were

identified for each DMP and DMR, which can probably be attributed to linkage disequilibrium which causes nearby loci to be jointly inherited. However, since some loci remained statistically differentially methylated after adjusting with the top meQTL, a combination of environmental and genetic factors could be at play, although the combined effect of multiple meQTLs, *trans*-meQTLs or other confounders cannot be excluded.

Investigating the association between DNA methylation and proximal gene expression gives insight to the potentially functionally relevant loci. Among the most statistically significant associations in the gene expression analysis is that of DNA methylation at cg00674440, located in the *MADD* gene body, and the expression of *MYBPC3* and *SPI1* transcripts. *MYBPC3* is a sarcomere component of heart muscle and mutations in the gene are associated with hypertrophic cardiomyopathy [102], while *SPI1* is a transcription factor with an ETS-domain and its expression in cardiac tissue was found to increase during myocardial infarction as a result of reduced DNA methylation, in mice [103]. A strong association was also identified between methylation at cg23194629, located in the gene body of *NUP88*, and *RABEP1* expression. *RABEP1* is involved in vesicular transport and autophagy [104] and one study has found its increased expression in heart and lung tissue to be associated pulmonary arterial hypertension in mice [105].

By looking at the association between methylation at DMPs and DMRs and cardiometabolic disease risk phenotypes, we can gain insight to the potential role of DNA methylation in the observed East/West discrepancy of CMDs. For this analysis, we also utilized the LURIC cohort since it consists of patients diagnosed with cardiometabolic diseases who have more severe phenotypes than the YFS cohort. Comparing the results from YFS and LURIC, DNA methylation at cg11472901, cg18948877, cg26740318, DMR\_11p15, DMR\_19q13, and DMR\_22q13 associated with various CMD-risk related phenotypes in both cohorts. Replication across both cohorts supports the hypothesis that altered methylation levels at these positions and regions may impact cardiometabolic health. Interestingly, associations observed for cg26740318,

DMR\_11p15 and DMR\_22q13 replicate closely between the two cohorts.

Increased methylation at cg26740318, located in the TSS1500 region of *LDHAL6A*, was associated with increased insulin treatment in YFS and decreased fasting insulin levels in LURIC. These findings were also similar for methylation at DMR\_11p15 which overlaps the cg26740318 locus. Higher methylation at both the DMP and DMR was found in East-originating Finns, which aligns with statistics that show that the incidence of type II diabetes is higher in north-eastern regions of Finland [106]. The *LDHAL6A* protein is a lactate dehydrogenase which is primarily expressed in the testes [107]; however, it was also found to be differentially expressed in pancreatic  $\beta$  cells of patients with type II diabetes [108].

Our data also revealed that increased DNA methylation at DMR\_22q13 associates with elevated triglyceride levels in both YFS and LURIC. Higher methylation levels in this region were found in East-originating Finns, again suggesting that they are metabolically worse off. Interestingly, a previous study identified an association between increased galactin-1 levels, encoded by *LSGAL1* whose expression was found to be associated with DMR\_22q13 methylation in this study, with increased triglycerides in humans [109]. Galectin-1 is a  $\beta$ -galactoside binding protein with functions in metabolism and inflammation and it has been linked to obesity and type II diabetes [110,111].

Out of the 21 DMPs identified between East and West-originating Finns, previous literature revealed association with CMD phenotypes at two other loci. DNA methylation at cg08946995 in whole blood was linked to coronary artery disease [112], while methylation at cg26207766 in liver biopsy samples was linked to obesity and type II diabetes [113]. These two sites did not associate with phenotypes in our main population analysis; however, this can partly be due to our DNA methylation data originating from blood rather than liver samples. A list of the main studies that have previously reported an association between any of the 21 DMPs and any phenotypes can be found in Table S9.

## Strength and limitations of the study

The multi-omic approach of this study is a considerable strength. Single-omic studies are easier to perform and interpret; however, they can provide limited insight into the mechanisms of complex diseases due to the interplay of different omics [114]. The bioinformatic analysis of multi-omic data allows for more accurate elucidation of the molecular mechanisms involved in CMD pathophysiology as well as the development of personalised treatment [115–117]. In our analysis, we integrated methylation, genome and transcriptome data, as well as blood measurements for CMD-related traits, for a more holistic investigation of the establishment and effect of DNA methylation. Our findings highlight the particular importance of taking the genetic component into consideration, particularly when studying a population with known genetic substructure, as we determined that all the DMPs and DMRs have a strong genetic basis. If genetic data were not available, online databases documenting known associations between DNA methylation and SNPs, such as the Genetics of DNA methylation consortium [118], can be used. The availability of genetic data in our case, however, also allowed us to account for the genetic factors in the transcriptomic and phenotypic analysis, in order to minimize their relative effect on the results and, as much as possible, isolate the environmental effects.

A significant concern related to DNA methylation measurements with available array technology is the poor reproducibility at specific probes [119–121]. The reliability of 10 probes out of the 21 sites identified in this study was included in a previously published reliability measurement study where the same samples were analysed with both 450K and EPIC arrays [119]. All 10 probes were found to have a reliability of  $\geq 75\%$ , which is considered highly reliable and therefore suggests our results may be reproducible. Replication in a different cohort, using different methodology and sequencing techniques, would be required to confirm our results.

## Conclusions

Taken together, the main results indicate that some of the differentially methylated loci between

East- and West-Finns may be functionally active in modulating transcription and may be contributing to the CMD difference between Eastern and Western Finns. Our analysis has exploited the availability of multi-omic data to account for the genetic effect so that our results will more accurately reflect the extrinsic, rather than intrinsic, influence on DNA methylation levels.

## Disclosure statement

No potential conflict of interest was reported by the author(s).

## Funding

JC was supported by funding from The Finnish Foundation for Cardiovascular Research, Yrjö Jahnsson Foundation and the Tampere City Science Fund. PPM was supported by the Academy of Finland (Grant number: 349708). The Young Finns Study has been financially supported by the Academy of Finland: grants 356405, 322098, 286284, 134309 (Eye), 126925, 121584, 124282, 129378 (Salve), 117797 (Gendi), and 141071 (Skidi); the Social Insurance Institution of Finland; Competitive State Research Financing of the Expert Responsibility area of Kuopio, Tampere and Turku University Hospitals (grant X51001); Juho Vainio Foundation; Paavo Nurmi Foundation; Finnish Foundation for Cardiovascular Research; Finnish Cultural Foundation; The Sigrid Juselius Foundation; Tampere Tuberculosis Foundation; Emil Aaltonen Foundation; Yrjö Jahnsson Foundation; Signe and Ane Gyllenberg Foundation; Diabetes Research Foundation of Finnish Diabetes Association; EU Horizon 2020 (grant 755320 for TAXINOMISIS and grant 848146 for To Aition); European Research Council (grant 742927 for MULTIEPIGEN project); Tampere University Hospital Supporting Foundation, Finnish Society of Clinical Chemistry, the Cancer Foundation Finland, and BETTER4U\_EU (Preventing obesity through Biologically and Behaviorally Tailored interventions for you; project number: 101080117). The genome-wide methylation analyses in the LURIC cohort were financially supported by the German Federal Ministry for Education and Research (BMBF; grant agreement numbers 01EA1808A and 01EA1411A) within the framework of the German Competence Cluster for Nutrition and Cardiovascular Health (nutriCARD) (Halle-Jena-Leipzig). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

## Author contributions

JC, SM and ER planned the study. JC, SM and ER designed the data analysis and JC performed all of the analyses. SR, PPM and LPL instructed in the data analysis, while MJ, PPM, LPL, NM, MEK, WM, TL and MK provided data. MK, OR,

TL and ER acquired funding and provided general leadership. JC wrote the first draft of the manuscript and edited it in response to the comments of the co-authors. SM, SR, DK and ER participated in the manuscript writing and all co-authors read and revised the final manuscript.

## Data availability statement

Data used in this study are considered sensitive due to potential identifiability of participants. Data sharing is therefore restricted under the regulations on professional secrecy (Act on the Openness of Government Activities, 612/1999) and on sensitive personal data (Personal Data Act, 523/1999, implementing the EU data protection directive 95/46/EC). Due to these legal restrictions, the data from this study cannot be stored in public repositories or otherwise made publicly available. The data are available from the authors upon reasonable request.

## Ethics statement

Written informed consent by all study participants was required in accordance with the Helsinki declaration and both studies were also assessed by respective ethical boards. YFS was approved by the 1st ethical committee of the Hospital District of Southwest Finland (21 September 2010) and by local ethical committees. LURIC was approved by the ethics committee of the 'Landesärztekammer RheinlandPfalz' state in Germany (LURIC,#83725597(1394)).

## ORCID

Joanna Ciantar  <http://orcid.org/0000-0002-3715-480X>

## References

- [1] Norio R. Finnish disease heritage II: population prehistory and genetic roots of finns. *Hum Genet.* 2003 May;112(5–6):457–469. doi: [10.1007/s00439-002-0876-2](https://doi.org/10.1007/s00439-002-0876-2)
- [2] Raento P, Husso K. Cultural diversity in Finland. *Fennia.* 2002 Jan;180(1–2):151–164.
- [3] Salmela E, Lappalainen T, Fransson I, et al. Genome-wide analysis of single nucleotide polymorphisms uncovers population structure in Northern Europe. Gemmell NJ, editor. *PLOS ONE.* [2008 Oct 24];3(10):e3519. doi: [10.1371/journal.pone.0003519](https://doi.org/10.1371/journal.pone.0003519)
- [4] Vähämurto L, Pahkala K, Magnussen CG, et al. Coronary heart disease risk factor levels in eastern and western Finland from 1980 to 2011 in the cardiovascular risk in young finns study. *Atherosclerosis.* 2019 Jan;280:92–98. doi: [10.1016/j.atherosclerosis.2018.11.007](https://doi.org/10.1016/j.atherosclerosis.2018.11.007)
- [5] Puska P, Salonen J, Nissinen A, et al. The North Karelia project. *Prev Med.* 1983 Jan;12(1):191–195. doi: [10.1016/0091-7435\(83\)90193-7](https://doi.org/10.1016/0091-7435(83)90193-7)
- [6] Juonala M, Viikari JSA, Hutri-Kähönen N, et al. The 21-year follow-up of the cardiovascular risk in young finns study: risk factor levels, secular trends and east–west difference. *J Intern Med.* 2004 Apr;255(4):457–468. doi: [10.1111/j.1365-2796.2004.01308.x](https://doi.org/10.1111/j.1365-2796.2004.01308.x)
- [7] Viikari J, Rönnemaa T, Seppänen A, et al. Serum lipids and lipoproteins in children, adolescents and young adults in 1980–1986. *Ann Med.* 1991 Jan;23(1):53–59. doi: [10.3109/07853899109147931](https://doi.org/10.3109/07853899109147931)
- [8] Raitakari O, Kivela A, Pahkala K, et al. Long-term tracking and population characteristics of lipoprotein (a) in the cardiovascular risk in young finns study. *Atherosclerosis.* 2022 Sep;356:18–27. doi: [10.1016/j.atherosclerosis.2022.07.009](https://doi.org/10.1016/j.atherosclerosis.2022.07.009)
- [9] Jousilahti P, Laatikainen T, Salomaa V, et al. 40-year CHD mortality trends and the role of risk factors in mortality decline: the north Karelia project experience. *Glob Heart.* [2016 Jun 1];11(2):207. doi: [10.1016/j.gheart.2016.04.004](https://doi.org/10.1016/j.gheart.2016.04.004)
- [10] Vartiainen E, Laatikainen T, Tapanainen H, et al. Changes in serum cholesterol and diet in North Karelia and all Finland. *Glob Heart.* [2016 Jun 1];11(2):179. doi: [10.1016/j.gheart.2016.04.006](https://doi.org/10.1016/j.gheart.2016.04.006)
- [11] National Health Index 2019–2021. Regional differences in morbidity and work disability are similar between wellbeing services counties: the best situation in Uusimaa and Ostrobothnia [internet]. Report No.: 74/2023. Available from THL; 2023 May. Available from: <https://urn.fi/URN:NBN:fi-fe202401091947>
- [12] THL's Morbidity Index 2019: Large regional differences in morbidity [Internet]. Report No.: 19/2022. Available from THL; 2022 May. Available from: <https://urn.fi/URN:NBN:fi-fe2022051936892>
- [13] Kerminen S, Martin AR, Koskela J, et al. Geographic variation and bias in the polygenic scores of complex diseases and traits in Finland. *Am J Hum Genet.* 2019 Jun;104(6):1169–1181. doi: [10.1016/j.ajhg.2019.05.001](https://doi.org/10.1016/j.ajhg.2019.05.001)
- [14] Tyynele P, Goebeler S, Ilveskoski E, et al. Age-dependent interaction of apolipoprotein E gene with eastern birthplace in Finland affects severity of coronary atherosclerosis and risk of fatal myocardial infarction—Helsinki sudden death study. *Ann Med.* 2013 May;45(3):213–219. doi: [10.3109/07853890.2012.727021](https://doi.org/10.3109/07853890.2012.727021)
- [15] THL's morbidity index 2000–2019. [Internet]. [cited 2024 Jan 10]. Available from: <https://terveytemme.fi/sairastavuusindeksi>
- [16] Jin B, Li Y, Robertson KD. DNA methylation: superior or subordinate in the epigenetic hierarchy? *Genes Cancer.* [2011 Jun 1];2(6):607–617. doi: [10.1177/1947601910393957](https://doi.org/10.1177/1947601910393957)
- [17] Unnikrishnan A, Freeman WM, Jackson J, et al. The role of DNA methylation in epigenetics of aging. *Pharmacol Ther.* 2019 Mar;195:172–185. doi: [10.1016/j.pharmthera.2018.11.001](https://doi.org/10.1016/j.pharmthera.2018.11.001)

- [18] Jones MJ, Goodman SJ, Kobor MS. DNA methylation and healthy human aging. *Aging Cell*. 2015 Dec;14(6):924–932. doi: [10.1111/accel.12349](https://doi.org/10.1111/accel.12349)
- [19] Chen C, Wang Z, Ding Y, et al. DNA methylation: from cancer biology to clinical perspectives. *Front Biosci-Landmark*. [2022 Dec 20];27(12):326. doi: [10.31083/j.fbl2712326](https://doi.org/10.31083/j.fbl2712326)
- [20] Nishiyama A, Nakanishi M. Navigating the DNA methylation landscape of cancer. *Trends Genet*. 2021 Nov;37(11):1012–1027. doi: [10.1016/j.tig.2021.05.002](https://doi.org/10.1016/j.tig.2021.05.002)
- [21] Rasmi Y, Shokati A, Hassan A, et al. The role of DNA methylation in progression of neurological disorders and neurodegenerative diseases as well as the prospect of using DNA methylation inhibitors as therapeutic agents for such disorders. *IBRO Neurosci Rep*. 2023 Jun;14:28–37. doi: [10.1016/j.ibneur.2022.12.002](https://doi.org/10.1016/j.ibneur.2022.12.002)
- [22] Younesian S, Yousefi AM, Momeny M, et al. The DNA methylation in neurological diseases. *Cells*. [2022 Oct 31];11(21):3439. doi: [10.3390/cells11213439](https://doi.org/10.3390/cells11213439)
- [23] Prasher D, Greenway SC, Singh RB. The impact of epigenetics on cardiovascular disease. *Biochem Cell Biol*. 2020 Feb;98(1):12–22. doi: [10.1139/bcb-2019-0045](https://doi.org/10.1139/bcb-2019-0045)
- [24] Shi Y, Zhang H, Huang S, et al. Epigenetic regulation in cardiovascular disease: mechanisms and advances in clinical trials. *Signal Transduct Target Ther*. [2022 Jun 25];7(1):200. doi: [10.1038/s41392-022-01055-2](https://doi.org/10.1038/s41392-022-01055-2)
- [25] Westerman KE, Ordovás JM. DNA methylation and incident cardiovascular disease. *Curr Opin Clin Nutr Metab Care*. 2020 Jul;23(4):236–240. doi: [10.1097/MCO.0000000000000659](https://doi.org/10.1097/MCO.0000000000000659)
- [26] Gomez-Alonso MDC, Kretschmer A, Wilson R, et al. DNA methylation and lipid metabolism: an EWAS of 226 metabolic measures. *Clin Epigenetics*. 2021 Dec;13(1):7. doi: [10.1186/s13148-020-00957-8](https://doi.org/10.1186/s13148-020-00957-8)
- [27] Christiansen C, Castillo-Fernandez JE, Domingo-Relloso A, et al. Novel DNA methylation signatures of tobacco smoking with trans-ethnic effects. *Clin Epigenetics*. 2021 Dec;13(1):36. doi: [10.1186/s13148-021-01018-4](https://doi.org/10.1186/s13148-021-01018-4)
- [28] Joehanes R, Just AC, Marioni RE, et al. Epigenetic signatures of cigarette smoking. *Circ Cardiovasc Genet*. 2016 Oct;9(5):436–447. doi: [10.1161/CIRCGENETICS.116.001506](https://doi.org/10.1161/CIRCGENETICS.116.001506)
- [29] Mishra PP, Hänninen I, Raitoharju E, et al. Epigenome-450K-wide methylation signatures of active cigarette smoking: the young finns study. *Biosci Rep*. [2020 Jul 31];40(7):BSR20200596. doi: [10.1042/BSR20200596](https://doi.org/10.1042/BSR20200596)
- [30] Huang RC, Garratt ES, Pan H, et al. Genome-wide methylation analysis identifies differentially methylated CpG loci associated with severe obesity in childhood. *Epigenetics*. [2015 Nov 2];10(11):995–1005. doi: [10.1080/15592294.2015.1080411](https://doi.org/10.1080/15592294.2015.1080411)
- [31] Demerath EW, Guan W, Grove ML, et al. Epigenome-wide association study (EWAS) of BMI, BMI change and waist circumference in African American adults identifies multiple replicated loci. *Hum Mol Genet*. [2015 Aug 1];24(15):4464–4479. doi: [10.1093/hmg/ddv161](https://doi.org/10.1093/hmg/ddv161)
- [32] Gaunt TR, Shihab HA, Hemani G, et al. Systematic identification of genetic influences on methylation across the human life course. *Genome Biol*. 2016 Dec;17(1):61. doi: [10.1186/s13059-016-0926-z](https://doi.org/10.1186/s13059-016-0926-z)
- [33] Min JL, Hemani G, Hannon E, et al. Genomic and phenotypic insights from an atlas of genetic effects on DNA methylation. *Nat Genet*. [2021 Sep 1];53(9):1311–1321. doi: [10.1038/s41588-021-00923-x](https://doi.org/10.1038/s41588-021-00923-x)
- [34] Hawe JS, Wilson R, Schmid KT, et al. Genetic variation influencing DNA methylation provides insights into molecular mechanisms regulating genomic function. *Nat Genet*. 2022 Jan;54(1):18–29. doi: [10.1038/s41588-021-00969-x](https://doi.org/10.1038/s41588-021-00969-x)
- [35] Husquin LT, Rotival M, Fagny M, et al. Exploring the genetic basis of human population differences in DNA methylation and their causal impact on immune gene regulation. *Genome Biol*. 2018 Dec;19(1):222. doi: [10.1186/s13059-018-1601-3](https://doi.org/10.1186/s13059-018-1601-3)
- [36] Benincasa G, Napoli C, DeMeo DL. Transgenerational epigenetic inheritance of cardiovascular diseases: a network medicine perspective. *Matern Child Health J*. 2024 Apr;28(4):617–630. doi: [10.1007/s10995-023-03886-z](https://doi.org/10.1007/s10995-023-03886-z)
- [37] Raitakari OT, Juonala M, Ronnema T, et al. Cohort profile: the cardiovascular risk in young finns study. *Int J Epidemiol*. [2008 Dec 1];37(6):1220–1226. doi: [10.1093/ije/dym225](https://doi.org/10.1093/ije/dym225)
- [38] Juonala M, Viikari JSA, Kähönen M, et al. Geographic origin as a determinant of carotid artery intima-media thickness and brachial artery flow-mediated dilation: the cardiovascular risk in young finns study. *Arterioscler Thromb Vasc Biol*. 2005 Feb;25(2):392–398. doi: [10.1161/01.ATV.0000150042.43866.7f](https://doi.org/10.1161/01.ATV.0000150042.43866.7f)
- [39] Winkelmann BR, März W, Boehm BO, et al. Rationale and design of the LURIC study - a resource for functional genomics, pharmacogenomics and long-term prognosis of cardiovascular disease. *Pharmacogenomics*. [2001 Feb 1];2(1s1):S1–73. doi: [10.1517/14622416.2.1.S1](https://doi.org/10.1517/14622416.2.1.S1)
- [40] Fortin JP, Triche TJ, Hansen KD, et al. Preprocessing, normalization and integration of the illumina HumanMethylationEPIC array with minfi. Hancock J, editor. *Bioinformatics*. [2017 Feb 15];33(4):558–560. doi: [10.1093/bioinformatics/btw691](https://doi.org/10.1093/bioinformatics/btw691)
- [41] Mishra PP, Mishra BH, Raitoharju E, et al. Gene set based integrated methylome and transcriptome analysis reveals potential molecular mechanisms linking

- cigarette smoking and related diseases. *Omics J Integr Biol.* [2023 May 1];27(5):193–204. doi: [10.1089/omi.2023.0028](https://doi.org/10.1089/omi.2023.0028)
- [42] Houseman EA, Accomando WP, Koestler DC, et al. DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinformatics.* 2012 Dec;13(1):86. doi: [10.1186/1471-2105-13-86](https://doi.org/10.1186/1471-2105-13-86)
- [43] Lehne B, Drong AW, Loh M, et al. A coherent approach for analysis of the illumina HumanMethylation450 BeadChip improves data quality and performance in epigenome-wide association studies. *Genome Biol.* 2015 Dec;16(1):37. doi: [10.1186/s13059-015-0600-x](https://doi.org/10.1186/s13059-015-0600-x)
- [44] Laaksonen J, Mishra PP, Seppälä I, et al. Mitochondrial genome-wide analysis of nuclear DNA methylation quantitative trait loci. *Hum Mol Genet.* [2022 May 19];31(10):1720–1732. doi: [10.1093/hmg/ddab339](https://doi.org/10.1093/hmg/ddab339)
- [45] Peters TJ, Buckley MJ, Statham AL, et al. De Novo identification of differentially methylated regions in the human genome. *Epigenetics Chromatin.* 2015 Dec;8(1):6. doi: [10.1186/1756-8935-8-6](https://doi.org/10.1186/1756-8935-8-6)
- [46] Lähteelä K, Kunnas T, Lyytikäinen LP, et al. No association of nineteen COX-2 gene variants to preclinical markers of atherosclerosis the cardiovascular risk in young finns study. *BMC Med Genet.* 2012 Dec;13(1):32. doi: [10.1186/1471-2350-13-32](https://doi.org/10.1186/1471-2350-13-32)
- [47] Taliun D, Harris DN, Kessler MD, et al. Sequencing of 53,831 diverse genomes from the NHLBI TOPMed program. *Nature.* [2021 Feb 11];590(7845):290–299. doi: [10.1038/s41586-021-03205-y](https://doi.org/10.1038/s41586-021-03205-y)
- [48] Kleber ME, Seppälä I, Pilz S, et al. Genome-wide association study identifies 3 genomic loci significantly associated with serum levels of homoarginine: the AtheroRemo consortium. *Circ Cardiovasc Genet.* 2013 Oct;6(5):505–513. doi: [10.1161/CIRCGENETICS.113.000108](https://doi.org/10.1161/CIRCGENETICS.113.000108)
- [49] Sudmant PH, Rausch T, Gardner EJ, et al. An integrated map of structural variation in 2,504 human genomes. *Nature.* [2015 Oct 1];526(7571):75–81. doi: [10.1038/nature15394](https://doi.org/10.1038/nature15394)
- [50] Elovainio M, Taipale T, Seppälä I, et al. Activated immune-inflammatory pathways are associated with long-standing depressive symptoms: evidence from gene-set enrichment analyses in the young finns study. *J Psychiatr Res.* 2015 Dec;71:120–125. doi: [10.1016/j.jpsychires.2015.09.017](https://doi.org/10.1016/j.jpsychires.2015.09.017)
- [51] Ritchie ME, Phipson B, Wu D, et al. Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* [2015 Apr 20];43(7):e47–e47. doi: [10.1093/nar/gkv007](https://doi.org/10.1093/nar/gkv007)
- [52] Raitoharju E, Seppälä I, Lyytikäinen LP, et al. Blood hsa-miR-122-5p and hsa-miR-885-5p levels associate with fatty liver and related lipoprotein metabolism—the young finns study. *Sci Rep.* [2016 Dec 5];6(1):38262. doi: [10.1038/srep38262](https://doi.org/10.1038/srep38262)
- [53] Belsky DW, Caspi A, Corcoran DL, et al. DunedinPACE, a DNA methylation biomarker of the pace of aging. *Elife.* [2022 Jan 14];11:e73420. doi: [10.7554/eLife.73420](https://doi.org/10.7554/eLife.73420)
- [54] Higgins-Chen AT, Thrush KL, Wang Y, et al. A computational solution for bolstering reliability of epigenetic clocks: implications for clinical trials and longitudinal tracking. *Nat Aging.* [2022 Jul 15];2(7):644–661. doi: [10.1038/s43587-022-00248-2](https://doi.org/10.1038/s43587-022-00248-2)
- [55] Chang CC, Chow CC, Tellier LC, et al. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience.* 2015 Dec;4(1):7. doi: [10.1186/s13742-015-0047-8](https://doi.org/10.1186/s13742-015-0047-8)
- [56] Smedley D, Haider S, Ballester B, et al. BioMart – biological queries made easy. *BMC Genomics.* 2009 Dec;10(1):22. doi: [10.1186/1471-2164-10-22](https://doi.org/10.1186/1471-2164-10-22)
- [57] Alberti KGMM, Eckel RH, Grundy SM, et al. Harmonizing the metabolic syndrome: a joint interim statement of the international diabetes federation task force on epidemiology and prevention. *Natl Heart Lung Blood Inst Am Heart Assoc World Heart Fed Int Atherosclerosis Soc Int Assoc Study Obes Circ.* [2009 Oct 20];120(16):1640–1645. doi: [10.1161/CIRCULATIONAHA.109.192644](https://doi.org/10.1161/CIRCULATIONAHA.109.192644)
- [58] Campagna MP, Xavier A, Lechner-Scott J, et al. Epigenome-wide association studies: current knowledge, strategies and recommendations. *Clin Epigenetics.* 2021 Dec;13(1):214. doi: [10.1186/s13148-021-01200-8](https://doi.org/10.1186/s13148-021-01200-8)
- [59] Kere J. Human population genetics: lessons from Finland. *Annu Rev Genomics Hum Genet.* 2001 Sep;2(1):103–128. doi: [10.1146/annurev.genom.2.1.103](https://doi.org/10.1146/annurev.genom.2.1.103)
- [60] Hannelius U, Salmela E, Lappalainen T, et al. Population substructure in Finland and Sweden revealed by the use of spatial coordinates and a small number of unlinked autosomal SNPs. *BMC Genet.* 2008 Dec;9(1):54. doi: [10.1186/1471-2156-9-54](https://doi.org/10.1186/1471-2156-9-54)
- [61] Jakkula E, Rehnström K, Varilo T, et al. The genome-wide patterns of variation expose significant substructure in a founder population. *Am J Hum Genet.* 2008 Dec;83(6):787–794. doi: [10.1016/j.ajhg.2008.11.005](https://doi.org/10.1016/j.ajhg.2008.11.005)
- [62] Kerminen S, Havulinna AS, Hellenthal G, et al. Fine-scale genetic structure in Finland. *G3 GenesGenomesgenetics.* [2017 Oct 1];7(10):3459–3468. doi: [10.1534/g3.117.300217](https://doi.org/10.1534/g3.117.300217)
- [63] Martin AR, Karczewski KJ, Kerminen S, et al. Haplotype sharing provides insights into fine-scale population history and disease in Finland. *Am J Hum Genet.* 2018 May;102(5):760–775. doi: [10.1016/j.ajhg.2018.03.003](https://doi.org/10.1016/j.ajhg.2018.03.003)

- [64] Karvonen JS, Rimpelä AH. Diminishing regional contrasts? The East–West-divide in health behaviour among Finnish adolescents. *Health Place*. 1998 Jun;4(2):161–170. doi: [10.1016/S1353-8292\(98\)00008-2](https://doi.org/10.1016/S1353-8292(98)00008-2)
- [65] Siirilä S, Vaattovaara M, Viljanen V. Well-being in Finland: a comparison of municipalities and residential differentiation in two cities. *Fennia*. 2002;180(1–2):141–149.
- [66] Fina S, Heider B, Mattila M, et al. Unequal Finland: regional socio-economic disparities in Finland. 2021 [cited 2024 Jan 22]. doi: [10.13140/RG.2.2.13328.12808](https://doi.org/10.13140/RG.2.2.13328.12808)
- [67] Bhatnagar A. Environmental determinants of cardiovascular disease. *Circ Res*. [2017 Jul 7];121(2):162–180. doi: [10.1161/CIRCRESAHA.117.306458](https://doi.org/10.1161/CIRCRESAHA.117.306458)
- [68] Schultz WM, Kelli HM, Lisko JC, et al. Socioeconomic status and cardiovascular outcomes: challenges and interventions. *Circulation*. [2018 May 15];137(20):2166–2178. doi: [10.1161/CIRCULATIONAHA.117.029652](https://doi.org/10.1161/CIRCULATIONAHA.117.029652)
- [69] Needham BL, Smith JA, Zhao W, et al. Life course socioeconomic status and DNA methylation in genes related to stress reactivity and inflammation: the multi-ethnic study of atherosclerosis. *Epigenetics*. [2015 Oct 3];10(10):958–969. doi: [10.1080/15592294.2015.1085139](https://doi.org/10.1080/15592294.2015.1085139)
- [70] Cronjé HT, Elliott HR, Nienaber-Rousseau C, et al. Leveraging the urban–rural divide for epigenetic research. *Epigenomics*. 2020 Jun;12(12):1071–1081. doi: [10.2217/epi-2020-0049](https://doi.org/10.2217/epi-2020-0049)
- [71] the BIOS Consortium, the Lifepath consortium, Fiorito G, McCrory C, et al. Socioeconomic position, lifestyle habits and biomarkers of epigenetic aging: a multi-cohort analysis. *Aging (Albany NY)*. [2019 Apr 14];11(7):2045–2070. doi: [10.18632/aging.101900](https://doi.org/10.18632/aging.101900)
- [72] Woodward M, Peters SAE, Batty GD, et al. Socioeconomic status in relation to cardiovascular disease and cause-specific mortality: a comparison of Asian and Australasian populations in a pooled analysis. *BMJ Open*. 2015 Mar;5(3):e006408. doi: [10.1136/bmjopen-2014-006408](https://doi.org/10.1136/bmjopen-2014-006408)
- [73] Okun S, Peek A, Igdoura SA. Neuraminidase 4 (NEU4): new biological and physiological player. *Glycobiology*. [2023 Apr 19];33(3):182–187. doi: [10.1093/glycob/cwad008](https://doi.org/10.1093/glycob/cwad008)
- [74] Carrizosa-Molina T, Casillas-Díaz N, Pérez-Nadador I, et al. Methylation analysis by targeted bisulfite sequencing in large for gestational age (LGA) newborns: the LARGAN cohort. *Clin Epigenetics*. [2023 Dec 13];15(1):191. doi: [10.1186/s13148-023-01612-8](https://doi.org/10.1186/s13148-023-01612-8)
- [75] Li S, Kim E, Wong EM, et al. Twin birth changes DNA methylation of subsequent siblings. *Sci Rep*. [2017 Aug 16];7(1):8463. doi: [10.1038/s41598-017-08595-6](https://doi.org/10.1038/s41598-017-08595-6)
- [76] Hara K, Maruki Y, Long X, et al. Raptor, a binding partner of target of rapamycin (TOR), mediates TOR action. *Cell*. 2002 Jul;110(2):177–189. doi: [10.1016/S0092-8674\(02\)00833-4](https://doi.org/10.1016/S0092-8674(02)00833-4)
- [77] Kim DH, Sarbassov DD, Ali SM, et al. mTOR interacts with raptor to form a nutrient-sensitive complex that signals to the cell growth machinery. *Cell*. 2002 Jul;110(2):163–175. doi: [10.1016/S0092-8674\(02\)00808-5](https://doi.org/10.1016/S0092-8674(02)00808-5)
- [78] Takahara T, Amemiya Y, Sugiyama R, et al. Amino acid-dependent control of mTORC1 signaling: a variety of regulatory modes. *J Biomed Sci*. 2020 Dec;27(1):87. doi: [10.1186/s12929-020-00679-2](https://doi.org/10.1186/s12929-020-00679-2)
- [79] Zhu Q, Qiao R, Di F, et al. Hypomethylation of RPTOR in peripheral blood is associated with very early-stage lung cancer. *Clin Chim Acta*. 2022 Dec;537:173–180. doi: [10.1016/j.cca.2022.10.014](https://doi.org/10.1016/j.cca.2022.10.014)
- [80] Tang Q, Holland-Letz T, Slynko A, et al. DNA methylation array analysis identifies breast cancer associated *RPTOR*, *MGRN1* and *RAPSN* hypomethylation in peripheral blood DNA. *Oncotarget*. [2016 Sep 27];7(39):64191–64202. doi: [10.18632/oncotarget.11640](https://doi.org/10.18632/oncotarget.11640)
- [81] Yin Y, Lei S, Li L, et al. RPTOR methylation in the peripheral blood and breast cancer in the Chinese population. *Genes Genomics*. 2022 Apr;44(4):435–443. doi: [10.1007/s13258-021-01182-0](https://doi.org/10.1007/s13258-021-01182-0)
- [82] Dugué PA, Milne RL, Southey MC. A prospective study of peripheral blood DNA methylation at *RPTOR*, *MGRN1* and *RAPSN* and risk of breast cancer. *Breast Cancer Res Treat*. 2017 Jan;161(1):181–183. doi: [10.1007/s10549-016-4032-4](https://doi.org/10.1007/s10549-016-4032-4)
- [83] Sciarretta S, Volpe M, Sadoshima J. Mammalian target of rapamycin signaling in cardiac physiology and disease. *Circ Res*. [2014 Jan 31];114(3):549–564. doi: [10.1161/CIRCRESAHA.114.302022](https://doi.org/10.1161/CIRCRESAHA.114.302022)
- [84] Zaina S, Heyn H, Carmona FJ, et al. DNA methylation map of human atherosclerosis. *Circ Cardiovasc Genet*. 2014 Oct;7(5):692–700. doi: [10.1161/CIRCGENETICS.113.000441](https://doi.org/10.1161/CIRCGENETICS.113.000441)
- [85] Juan AM, Bartolomei MS. Evolving imprinting control regions: KRAB zinc fingers hold the key. *Genes Dev*. [2019 Jan 1];33(1–2):1–3. doi: [10.1101/gad.322990.118](https://doi.org/10.1101/gad.322990.118)
- [86] Takahashi N, Gray D, Strogantsev R, et al. *ZFP57* and the targeted maintenance of postfertilization genomic imprints. *Cold Spring Harb Symp Quant Biol*. 2015;80:177–187. doi: [10.1101/sqb.2015.80.027466](https://doi.org/10.1101/sqb.2015.80.027466)
- [87] Jiang W, Shi J, Zhao J, et al. *ZFP57* dictates allelic expression switch of target imprinted genes. *Proc Natl Acad Sci*. [2021 Feb 2];118(5):e2005377118. doi: [10.1073/pnas.2005377118](https://doi.org/10.1073/pnas.2005377118)
- [88] Amarasekera M, Martino D, Ashley S, et al. Genome-wide DNA methylation profiling identifies a folate-sensitive region of differential methylation upstream of *ZFP57* -imprinting regulator in humans. *Faseb J*. 2014 Sep;28(9):4068–4076. doi: [10.1096/fj.13-249029](https://doi.org/10.1096/fj.13-249029)
- [89] Yusipov I, Bacalini MG, Kalyakulina A, et al. Age-related DNA methylation changes are sex-specific: a comprehensive assessment. *Aging (Albany NY)*. [2020 Dec 3];12(23):24057–24080. doi: [10.18632/aging.202251](https://doi.org/10.18632/aging.202251)

- [90] Laity JH, Lee BM, Wright PE. Zinc finger proteins: new insights into structural and functional diversity. *Curr Opin Struct Biol.* 2001 Feb;11(1):39–46. doi: [10.1016/S0959-440X\(00\)00167-6](https://doi.org/10.1016/S0959-440X(00)00167-6)
- [91] Razzaghi H, Santorico SA, Kamboh MI. Population-based resequencing of LIPG and ZNF202 genes in subjects with extreme HDL levels. *Front Genet* [Internet]. 2012 [cited 2024 Feb 5];3. Available from <http://journal.frontiersin.org/article/10.3389/fgene.2012.00089/abstract>. doi: [10.3389/fgene.2012.00089](https://doi.org/10.3389/fgene.2012.00089)
- [92] Vrins CLJ, Out R, Van Santbrink P, et al. Znf202 affects high density lipoprotein cholesterol levels and promotes hepatosteatosis in hyperlipidemic mice. Mohanraj R, editor. *PLOS ONE.* [2013 Feb 28];8(2):e57492. doi: [10.1371/journal.pone.0057492](https://doi.org/10.1371/journal.pone.0057492)
- [93] Senée V, Chelala C, Duchatelet S, et al. Mutations in GLIS3 are responsible for a rare syndrome with neonatal diabetes mellitus and congenital hypothyroidism. *Nat Genet.* 2006 Jun;38(6):682–687. doi: [10.1038/ng1802](https://doi.org/10.1038/ng1802)
- [94] Scherneck S, Nestler M, Vogel H, et al. Positional cloning of zinc finger domain transcription factor Zfp69, a candidate gene for obesity-associated diabetes contributed by mouse locus Nidd/sjl. Flint J, editor. *PloS Genet.* [2009 Jul 3];5(7):e1000541. doi: [10.1371/journal.pgen.1000541](https://doi.org/10.1371/journal.pgen.1000541)
- [95] Baglivo I, Esposito S, De Cesare L, et al. Genetic and epigenetic mutations affect the DNA binding capability of human ZFP57 in transient neonatal diabetes type 1. *FEBS Lett.* [2013 May 21];587(10):1474–1481. doi: [10.1016/j.febslet.2013.02.045](https://doi.org/10.1016/j.febslet.2013.02.045)
- [96] Fan D, Pang S, Chen J, et al. Identification and functional study of GATA4 gene regulatory variants in atrial septal defects. *BMC Cardiovasc Disord.* 2021 Dec;21(1):321. doi: [10.1186/s12872-021-02136-w](https://doi.org/10.1186/s12872-021-02136-w)
- [97] Kodo K, Nishizawa T, Furutani M, et al. GATA6 mutations cause human cardiac outflow tract defects by disrupting semaphorin-plexin signaling. *Proc Natl Acad Sci.* [2009 Aug 18];106(33):13933–13938. doi: [10.1073/pnas.0904744106](https://doi.org/10.1073/pnas.0904744106)
- [98] Jian H, Poetsch A. CASZ1: current implications in cardiovascular diseases and cancers. *Biomedicines.* [2023 Jul 24];11(7):2079. doi: [10.3390/biomedicines11072079](https://doi.org/10.3390/biomedicines11072079)
- [99] Belsky DW, Caspi A, Houts R, et al. Quantification of biological aging in young adults. *Proc Natl Acad Sci USA.* 2015 Jul 28 [cited 2024 Jan 16];112(30). Available from doi: [10.1073/pnas.1506264112](https://doi.org/10.1073/pnas.1506264112)
- [100] Elliott ML, Belsky DW, Knodt AR, et al. Brain-age in midlife is associated with accelerated biological aging and cognitive decline in a longitudinal birth cohort. *Mol Psychiatry.* 2021 Aug;26(8):3829–3838. doi: [10.1038/s41380-019-0626-7](https://doi.org/10.1038/s41380-019-0626-7)
- [101] Villicaña S, Bell JT. Genetic impacts on DNA methylation: research findings and future perspectives. *Genome Biol.* [2021 Apr 30];22(1):127. doi: [10.1186/s13059-021-02347-6](https://doi.org/10.1186/s13059-021-02347-6)
- [102] Tudurachi BS, Zăvoi A, Leonte A, et al. An update on MYBPC3 gene mutation in hypertrophic cardiomyopathy. *Int J Mol Sci.* [2023 Jun 22];24(13):10510. doi: [10.3390/ijms241310510](https://doi.org/10.3390/ijms241310510)
- [103] Liu Z, Shuai H. Upregulation of SPI1 during myocardial infarction aggravates cardiac tissue injury and disease progression through activation of the TLR4/NFκB axis. *Am J Transl Res.* [2022 Apr 15];14(4):2709–2727.
- [104] Millarte V, Spiess M. RABEP1/Rabaptin5: a link between autophagy and early endosome homeostasis. *Autophagy.* [2022 Mar 4];18(3):698–699. doi: [10.1080/15548627.2021.2021497](https://doi.org/10.1080/15548627.2021.2021497)
- [105] Gong Q, Hu Z, Jin Q, et al. Identification of JPX-RABEP1 pair as an immune-related biomarker and therapeutic target in pulmonary arterial hypertension by bioinformatics and experimental analyses. *Int J Mol Sci.* [2022 Dec 8];23(24):15559. doi: [10.3390/ijms232415559](https://doi.org/10.3390/ijms232415559)
- [106] Arffman M, Ilanne-Parikka P, Keskimäki I, et al. FinDM database on diabetes in Finland [Internet]. 2020. Available from: <http://urn.fi/URN:ISBN:978-952-343-492-9>
- [107] Chen X, Gu X, Shan Y, et al. Identification of a novel human lactate dehydrogenase gene LDHAL6A, which activates transcriptional activities of AP1(PMA). *Mol Biol Rep.* 2009 Apr;36(4):669–676. doi: [10.1007/s11033-008-9227-2](https://doi.org/10.1007/s11033-008-9227-2)
- [108] Ding L, Fan L, Xu X, et al. Identification of core genes and pathways in type 2 diabetes mellitus by bioinformatics analysis. *Mol Med Rep* [Internet]. 2019 Jul 24 [cited 2024 Jan 29]. doi: [10.3892/mmr.2019.10522](https://doi.org/10.3892/mmr.2019.10522)
- [109] Fryk E, Strindberg L, Lundqvist A, et al. Galectin-1 is inversely associated with type 2 diabetes independently of obesity – a SCAPIS pilot study. *Metab Open.* 2019 Dec;4:100017. doi: [10.1016/j.metop.2019.100017](https://doi.org/10.1016/j.metop.2019.100017)
- [110] Fryk E, Sundelin JP, Strindberg L, et al. Microdialysis and proteomics of subcutaneous interstitial fluid reveals increased galectin-1 in type 2 diabetes patients. *Metabolism.* 2016 Jul;65(7):998–1006. doi: [10.1016/j.metabol.2016.04.003](https://doi.org/10.1016/j.metabol.2016.04.003)
- [111] Liu X, Feng Q, Chen Y, et al. Proteomics-based identification of differentially-expressed proteins including galectin-1 in the blood plasma of type 2 diabetic patients. *J Proteome Res.* [2009 Mar 6];8(3):1255–1262. doi: [10.1021/pr800850a](https://doi.org/10.1021/pr800850a)
- [112] Uddin MDM, Nguyen NQH, Yu B, et al. Clonal hematopoiesis of indeterminate potential, DNA methylation, and risk for coronary artery disease. *Nat Commun.* [2022 Sep 12];13(1):5350. doi: [10.1038/s41467-022-33093-3](https://doi.org/10.1038/s41467-022-33093-3)
- [113] Kirchner H, Sinha I, Gao H, et al. Altered DNA methylation of glycolytic and lipogenic genes in liver from obese

- and type 2 diabetic patients. *Mol Metab.* 2016 Mar;5(3):171–183. doi: [10.1016/j.molmet.2015.12.004](https://doi.org/10.1016/j.molmet.2015.12.004)
- [114] Sun YV, Hu YJ. Integrative analysis of multi-omics data for discovery and functional studies of complex human diseases. In: *Advances in genetics* [internet]. Elsevier; 2016 [cited 2024 Jan 22]. p. 147–190. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S0065266015000516>
- [115] Leon-Mimila P, Wang J, Huertas-Vazquez A. Relevance of multi-omics studies in cardiovascular diseases. *Front Cardiovasc Med.* [2019 Jul 17];6:91. doi: [10.3389/fcvm.2019.00091](https://doi.org/10.3389/fcvm.2019.00091)
- [116] Sopic M, Vilne B, Gerdtts E, et al. Multiomics tools for improved atherosclerotic cardiovascular disease management. *Trends Mol Med.* 2023 Dec;29(12):983–995. doi: [10.1016/j.molmed.2023.09.004](https://doi.org/10.1016/j.molmed.2023.09.004)
- [117] Benincasa G, Suades R, Padró T, et al. Bioinformatic platforms for clinical stratification of natural history of atherosclerotic cardiovascular diseases. *Eur Heart J Cardiovasc Pharmacother.* [2023 Dec 14];9(8):758–769. doi: [10.1093/ehjcvp/pvad059](https://doi.org/10.1093/ehjcvp/pvad059)
- [118] GoDMC. Available from: <http://mqtl.db.godmc.org.uk/>
- [119] Sugden K, Hannon EJ, Arseneault L, et al. Patterns of reliability: assessing the reproducibility and integrity of DNA methylation measurement. *Patterns.* 2020 May;1(2):100014. doi: [10.1016/j.patter.2020.100014](https://doi.org/10.1016/j.patter.2020.100014)
- [120] 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 Dec;15(1):312. doi: [10.1186/1471-2105-15-312](https://doi.org/10.1186/1471-2105-15-312)
- [121] Dugué PA, English DR, MacInnis RJ, et al. Reliability of DNA methylation measures from dried blood spots and mononuclear cells using the HumanMethylation450k BeadArray. *Sci Rep.* [2016 Jul 26];6(1):30317. doi: [10.1038/srep30317](https://doi.org/10.1038/srep30317)