DNA methylation, cardiovascular risk and myocardial infarction: an epigenome-wide approach

PhD Candidate: Alba Fernández Sanlés

DOCTORAL THESIS UPF / 2019

Thesis supervisor:

Dr. Roberto Elosua Llanos

DEPARTMENT OF EXPERIMENTAL AND HEALTH SCIENCES

PhD Programme in Biomedicine

Universitat Pompeu Fabra Barcelona

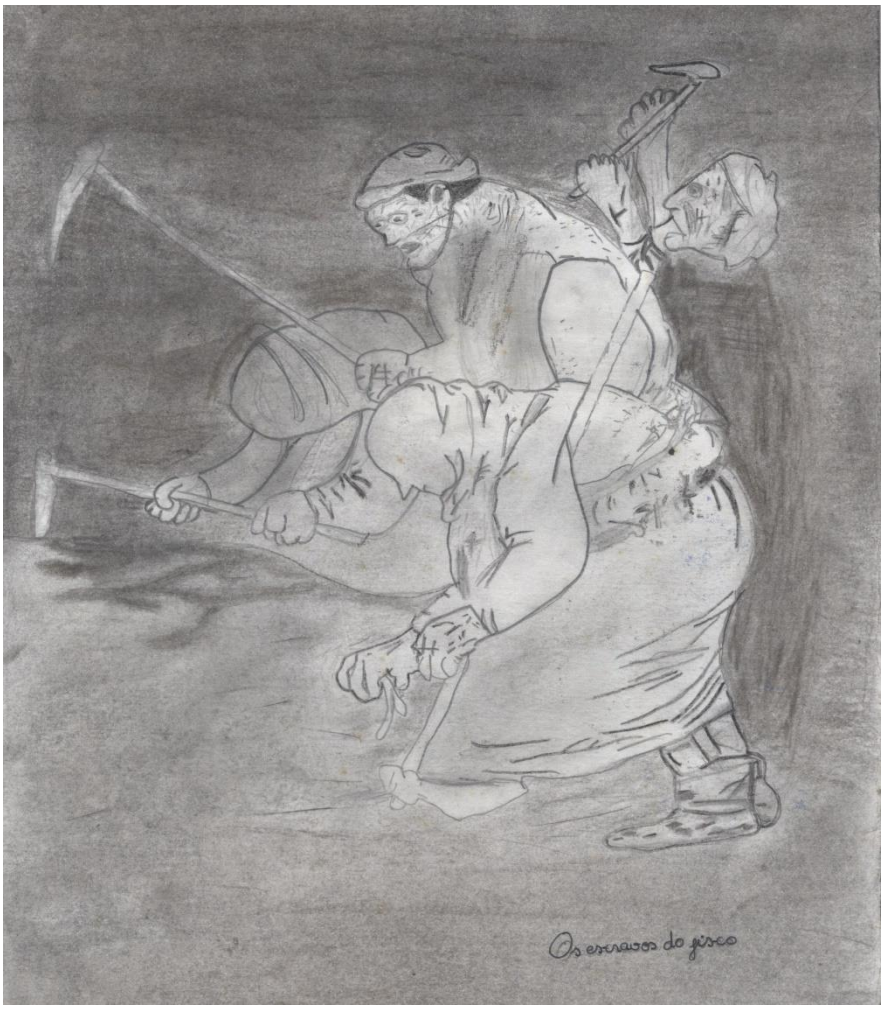
IMIM Parc de Salut MAR Barcelona

A mis padres. Por todo.

A Javi. Por sumar.

"We do not inherit the Earth from our ancestors; we borrow it from our children."

Antoine de Saint-Exupéry (1900-1944).



La mujer sin rostro podría haber respondido a Rosa. O a María, o a Joaquina, o a Cipriana. Él, a Pepe. Podrían haber cambiado las azadas por las redes de pesca y responder a Manuel. Podrían haberlas cambiado por los libros, pero les tocó nacer entonces. Entonces fue una época anterior a la propia de sus inquietudes. Ahora esas azadas son este libro.

Charcoal copy of "Os escravos do fisco" (Castelao, Álbum Nós), AF-S, 1999.

Agradecimientos

Que haya llegado el momento de escribir este apartado de mi tesis doctoral significa que poco más queda que imprimirla y luego defenderla. Y que esto sea posible se explica por la sinergia de múltiples factores, algunos en forma de actores concretos.

Roberto, uno de los dos actores principales. Aunque hubiese un algoritmo con una alta sensibilidad para buscar buenos supervisores de tesis, no podría haber dado con uno mejor. Gracias por haberme aceptado sin prejuicios por cambiarme de grupo y por haber confiado en mí para trabajar con autonomía supervisada. Por tu perfil docente, por tu perfil científico, por tu perfil de mentor, por tu perfil de líder y por tu perfil más humano. He aprendido mucho contigo y has dejado el listón muy alto para mis futuros jefes.

Sergi, mi hermano mayor de doctorado desde el día que volvió de su estancia en Alabama y se encontró con que tenía una hermana inesperada que venía de hacer laboratorio de biología molecular y celular. Gracias por tu paciencia con mis preguntas de código, por tus enseñanzas para poder contestar a los objetivos de esta tesis y por todos tus consejos.

Albert Prats, que me ha ayudado con dudas informáticas, e Isaac, Joan y Silvia, que lo han hecho con dudas estadísticas y de R. Los informáticos que dan soporte a la investigación con grandes bases de datos. Elaine y Gavin, los editores de los artículos y de esta tesis, respectivamente.

El grupo EGEC/REGICOR/CARIN, con una gran calidad científica y humana. Esmeralda, que hace más fácil todo lo administrativo. Leny, Dani, Gemma y Sònia, que se han encargado de los pasos previos a mi trabajo. Los voluntarios de las cohortes de esta tesis, imprescindibles. Los otros predocs y allegados (Albert, Ruwaida, Álvaro, Dolors, Ale, Camille, Jorge, María, Gabi, etc.), que somos una gran familia y nos cuidamos como tal en el día a día.

El otro actor con papel principal en esta tesis soy yo con mis circunstancias. La curiosidad, la perseverancia y la consciencia han sido el combustible para la motivación que conduce a los sueños. Pero sueños tenemos todos, a diferencia de un contexto afortunado. Yo estoy agradecida de haberlo tenido, empezando por mi contexto familiar.

Mis padres, las piezas centrales de este puzle. No hay que olvidar que gracias a los padres la mayoría de los investigadores predoctorales hemos encontrado el apoyo económico y moral que se necesitan, desde el momento de decidir qué estudiar, y que van *in crescendo* conforme avanza la carrera académica. Papá, tú no pudiste ir a la Universidad porque tuviste que hacerte

cargo de tu madre y de tu hermana al quedarte huérfano, pero eres un hombre muy culto que nada contracorriente en tierra firme, que siempre me ha inspirado. Mamá, tú tampoco pudiste estudiar para ser profesora, pronto te vinieron muchas cosas encima, y las cosas de ser mujer de aldea gallega en los setenta, pero personificas la energía renovable, la fortaleza basada en el coraje y la bondad desinteresada, lo que admiro por encima de todo. Los dos sois más biólogos que yo y mis compañeros que no sabemos qué planta o qué pájaro tenemos delante, y vosotros y abuela (y mis hermanos peludos Prim, Dora, Nieve, Lúa, Perla y Brisa, demás mascotas, Disney y Cía., y los libros que poníais a mi alcance) habéis determinado mi curiosidad y respeto por la naturaleza. Y todo lo que me habéis dado luego lo sabemos los tres. Por todo ello, soy la primera persona en la familia que pudo (no sólo quiso) estudiar una carrera (la elegida), y la primera mujer en ella que ha podido trabajar en lo que le gusta. Sé que si abuela no tuviese demencia, estaría orgullosa. Gracias por este regalo tan valioso.

Algunos de mis profesores, que han contribuido a quién soy y así, indirectamente, a esta tesis: Marga, Concha Paz, María José, Pili Pereiro, Mónica Paula, *Toxo. Cando a docencia é vocacional, nótase; milleiros de graciñas*.

Javi, que no ha parado de sumar, de ser fuente de inspiración, comprensión y apoyo desde que nos conocemos. Gracias por tu infinita paciencia y por confiar más en mí que yo misma. Por ser lo suficientemente parecido y lo necesariamente diferente a mí para complementarnos. Por dibujar en el mismo *render* tu futuro con mi idea abstracta del mío. Nuestros bichos, Trasno y Allo, que me han enseñado lo que es responsabilizarme, sobre todo Trasno, que me dio extracto de amor puro para sentirme fuerte y seguir.

Continuando con el seguir, Alfredo Berzal y Cristina Romero, que me enseñaron tanto durante mi año en Granada y me apoyaron y animaron a no tirar la toalla año y medio más tarde.

Miry, mi terapeuta y coach. Año tras año se dice que los doctorandos sufrimos más de depresión, ansiedad o estrés que el resto de la población general. Sin embargo, nadie conoce a nadie con alguno de estos trastornos y así poco se puede hacer para desnormalizar esta situación, pasando antes por normalizarla. Decidí reconocer este aspecto y ponerlo en la misma mesa que mi formación científica y técnica antes de llegar al equipo de Roberto y empezar esta tesis. Si he podido enfrentarme a mis miedos e inseguridades para conseguir mis metas ha sido porque también he aprendido a cuidarme, con mucho trabajo individual guiado por otra gran supervisora.

Barcelona, 15 de septiembre de 2019
Alba Fernández Sanlés

“The noblest pleasure is the joy of understanding.”

Leonardo da Vinci (1452-1519).

TABLE OF CONTENTS

Abbreviations and acronyms

450K: Human Methylation450 BeadChip
5mC: 5-methylCytosine
ACS: Acute Coronary Syndrome
AMI: Acute Myocardial Infarction
AMPK: AMP-activated Protein Kinase
ARIC: Atherosclerosis Risk in Communities
AUC: Area Under the Receiver Operating Characteristic Curve
BMI: Body Mass Index
BP: Blood Pressure
bp: base pair
CARDIoGRAMplusC4D: Coronary ARtery DIsease Genome-wide Replication And Meta-analysis plus Coronary Artery Disease Genetics Consortium
CGI: CpG Island
CHD: Coronary Heart Disease
circRNA: circular RNA
CpG: Cytosine-phosphate-Guanine
CVD: CardioVascular Disease
CVR: CardioVascular Risk
CVRF: CardioVascular Risk Factor
DALY: Disability-Adjusted Life-Year
dbGaP: database of Genotypes and Phenotypes
DBP: Dyastolic Blood Pressure
DNA: DeoxyriboNucleic Acid
DNMT: DNA MethylTransferase
EPIC: HumanMethylationEPIC BeadChip
EWAS: Epigenome-Wide Association Study
EXaC: EXome aggregation Consortium
FDR: False Discovery Rate
FHS: Framingham Heart Study
FOS: Framingham Offspring Study
FRESCO: *Función de Riesgo ESpañola de acontecimientos Coronarios y Otros*
GRS: Genetic Risk Score
GIV: Genetic Instrumental Variable
gnomAD: genome Aggregation Database
GWAS: Genome-Wide Association Studies
HapMap: Haplotype Map
HDL: High Density Lipoprotein

HDL-C: High Density Lipoprotein-Cholesterol
HF: Heart Failure
HPLC: High-Performance Liquid Chromatography
IDI: Integrated Discrimination Index
IDL: Intermediate Density Lipoprotein
LDL: Low Density Lipoprotein
LDL-C: Low Density Lipoprotein-Cholesterol
LINE: Long Interspersed Nuclear Element
lncRNA: long ncRNA
LPA: Light-intensity aerobic Physical Activity
mCH: non-CpG methylation
meQTL: methylation Quantitative Trait Loci
MI: Myocardial Infarction
MR: Mendelian Randomisation
miRNA: micro RNA
mRNA: messenger RNA
MRS: Methylation Risk Score
MESA: Multi-Ethnic Study of Atherosclerosis
METs: METabolic equivalents
MSP: Methylation-Specific PCR
MPA: Moderate-intensity aerobic Physical Activity
MVPA: Moderate- and Vigorous-intensity aerobic Physical Activity
NA: Not Available
NCD: Non-Communicable Disease
ncRNA: non-coding RNA
NGS: Next-Generation Sequencing
NRI: Net Reclassification Index
PA: Physical Activity
PCR: Polymorphism Chain Reaction
piRNA: piwi-interacting RNA
PMID: PubMed Unique Identifier
Q-Q plot: Quantile-Quantile plot
REGICOR: *REgistre GIroní del COR*
rRNA: ribosomal RNA
RNA: RiboNucleic Acid
RNAi: RNA interference
RRBS: Reduced Representation Bisulphite Sequencing
RVAS: Rare Variant Association Study
SBP: Systolic Blood Pressure
SCORE: Systematic COronary Risk Evaluation
SDI: Socio-Demographic Index

siRNA: small-interfering RNA
SNP: Single Nucleotide Polymorphism
STREGA: STrengthening the REporting of Genetic Association studies
T2D: Type 2 Diabetes
TG: TriGlycerides
TPA: total aerobic Physical Activity
tRNA: transfer RNA
VLDL: Very Low Density Lipoprotein.
VPA: Vigorous-intensity aerobic Physical Activity
VRF: Vascular Risk Factor
WES: Whole Exome Sequencing
WGS: Whole Genome Sequencing
WGBS: Whole Genome Bisulfite Sequencing
WHI: Women's Health Initiative
Xist: X-inactive specific transcript
YLL: Year of Life Lost
YLD: Year Lost due to Disability

Abstract

Cardiovascular diseases are a major challenge for public health. The complex biological networks underlying their etiology remain to be uncovered. Further studies that identify new actors and mediators are required to enable the design of more precise prevention strategies. DNA methylation is a mechanism regulating gene expression, and it is linked to complex diseases and related traits. Therefore, it is a potential source of biomarkers of cardiovascular disease.

We have tackled the associations of DNA methylation with cardiovascular diseases and related traits in population-based observational studies. We analysed the association between this mechanism in peripheral blood cells and both cardiovascular disease related exposures and outcomes, and further investigated the clinical relevance of some of those findings. For this purpose, we conducted epigenome-wide and candidate loci association studies, evaluated their predictive value for future cardiovascular risk and used Mendelian Randomisation to infer causality. We used three observational cohorts: *REgistre GIroní del COR*, Framingham Offspring Study and Women's Health Initiative.

We identified a non-linear dose-response to physical activity of two methylation sites. We found seven loci showing hypomethylation related to the cardiovascular risk factors load, excluding age. We reported 17 methylation sites associated with prevalent coronary heart disease or incident cardiovascular disease, some of them being novel. Those 17 biomarkers did not show an added predictive value over the established risk factors included in the current cardiovascular risk prediction functions. Finally, we could not infer the causality of their relationships with cardiovascular disease using a Mendelian Randomisation approach.

As a general conclusion, DNA methylation is associated with cardiovascular disease related traits and outcomes, although its role in the underlying biological networks is complex and remains to be explored in further studies.

Resumen

Las enfermedades cardiovasculares son uno de los principales retos para la Salud Pública. Las complejas redes moleculares subyacentes a su etiología no han sido esclarecidas. Son necesarios más estudios para identificar nuevos actores y mediadores que permitan diseñar estrategias preventivas más precisas. La metilación del ADN es un mecanismo que regula la expresión génica vinculada a enfermedades complejas y factores relacionados. Así, se presenta como una potencial fuente de biomarcadores de enfermedad cardiovascular.

Hemos estudiado la asociación entre la metilación del ADN y fenotipos cardiovasculares y relacionados en estudios observacionales en población general. Hemos analizado la asociación de este mecanismo en células de sangre periférica con factores de exposición relacionados con enfermedades cardiovasculares y con eventos cardiovasculares, así como la relevancia clínica de algunos de estos hallazgos. Para ello, hemos realizado estudios de asociación de epigenoma completo y de *loci* candidato, hemos evaluado su valor predictivo de riesgo cardiovascular y hemos utilizado la estrategia de aleatorización mendeliana para inferir causalidad. Hemos incluido tres cohortes observacionales: *REgistre GIroní del COR*, *Framingham Offspring Study* y *Women's Health Initiative*.

Hemos identificado una dosis-respuesta no lineal a la actividad física en dos sitios de metilación. Hemos encontrado siete *loci* con menores valores de metilación en relación a la acumulación de factores de riesgo independiente de la edad. Hemos descrito 17 sitios de metilación asociados a la prevalencia de enfermedad coronaria y la incidencia de enfermedad cardiovascular, algunos por primera vez. Estos 17 biomarcadores no aportaron un valor predictivo añadido sobre los factores de riesgo que se incluyen en las actuales funciones de predicción de riesgo cardiovascular. Por último, no hemos podido inferir causalidad utilizando la estrategia de aleatorización mendeliana.

Como conclusión general, la metilación del ADN se asocia a fenotipos relacionados con las enfermedades cardiovasculares y a eventos, pero su papel en las redes moleculares subyacentes es complejo y se necesitan más estudios para continuar avanzando en esta línea.

Resumo

As enfermidades cardiovasculares son un dos principais retos de Saúde Pública. As complexas redes moleculares subxacentes á súa etioloxía non están claras. Son necesarios máis estudos para identificar novos actores e mediadores que permitan deseñar estratexias preventivas máis precisas. A metilación do ADN é un mecanismo que regula a expresión xénica vencellado a enfermidades complexas e factores relacionados. Porén, preséntase coma unha potencial fonte de biomarcadores de enfermidade cardiovascular.

Estudamos a asociación entre a metilación do ADN e fenotipos cardiovasculares e relacionados en estudos observacionais en poboación xeral. Analizamos a asociación deste mecanismo en células de sangue periférico con factores de exposición relacionados con enfermidades cardiovasculares e con eventos cardiovasculares, así coma a relevancia clínica dalgúns destes achados. Para isto, realizamos estudos de asociación de epixenoma completo e de *loci* candidato, avaliamos o seu valor predictivo de risco cardiovascular e utilizamos a estratexia da aleatorización mendeliana para inferir causalidade. Incluímos tres cohortes observacionais: *REgistre GIroní del COR*, *Framingham Offspring Study* e *Women's Health Initiative*.

Identificamos unha dose-resposta non lineal á actividade física en dous sitios de metilación. Atopamos sete *loci* con menores valores de metilación en relación á acumulación de factores de risco independentes da idade. Describimos 17 sitios de metilación asociados á prevalencia de enfermidade coronaria e á incidencia de enfermidade cardiovascular, algúns por primeira vez. Estes 17 biomarcadores non aportaron un valor predictivo engadido sobre os factores de risco que se inclúen nas funcións de predición de risco cardiovascular actuais. Por último, non puidemos inferir causalidade utilizando a estratexia da aleatorización mendeliana.

Como conclusión xeral, a metilación do ADN asóciase a fenotipos relacionados coas enfermidades cardiovasculares e a eventos, pero o seu papel nas redes moleculares subxacentes é complexo e máis estudos son necesarios para avanzar nesta liña.

Resum

Les malalties cardiovasculars són un dels principals reptes de salut pública. Les complexes xarxes moleculars subjacents no han estat totalment definides. Per tant, es fan indispensables més estudis per a identificar nous actors i mediadors que permetin dissenyar estratègies preventives més precises. La metilació de l'ADN és un mecanisme que regula l'expressió gènica vinculat a malalties complexes i factors relacionats. Així, es presenta com a una potencial font de biomarcadors de malalties cardiovasculars.

Hem estudiat l'associació entre la metilació de l'ADN i fenotips cardiovasculars i relacionats en estudis observacionals en població general. Hem analitzat l'associació entre la metilació de l'ADN en cèl·lules de sang perifèrica amb factors d'exposició relacionats amb malalties cardiovasculars i amb esdeveniments cardiovasculars, així com la rellevància clínica d'algunes d'aquestes troballes. Per a això, hem realitzat estudis d'associació de epigenoma complet i de *loci* candidat, hem avaluat el seu valor predictiu de risc cardiovascular i hem utilitzat l'estratègia d'aleatorització mendeliana per inferir causalitat. Hem inclòs tres cohorts observacionals: Registre Gironí del COR, *Framingham Offspring Study* i *Women's Health Initiative*.

Hem identificat una dosi-resposta no lineal a l'activitat física en dos llocs de metilació. Hem trobat set *loci* amb menors valors de metilació amb relació a l'acumulació de factors de risc independent de l'edat. Hem descrit 17 llocs de metilació relacionats amb prevalença de malaltia coronària i incidència de malaltia cardiovascular, alguns per primera vegada. Aquests 17 biomarcadors no van aportar un valor predictiu afegit sobre els factors de risc que s'inclouen en les actuals funcions predictives de risc cardiovascular. Finalment, no hem pogut inferir causalitat utilitzant l'estratègia d'aleatorització mendeliana.

Com a conclusió general, la metilació de l'ADN s'associa a fenotips relacionats amb malalties cardiovasculars i a esdeveniments, però el seu paper en les xarxes moleculars subjacents és complex i es necessiten més estudis per continuar avançant en aquesta línia.

1. INTRODUCTION

Illustration by Gérard Dubois

"Dende aquí vexo un camiño
que non sei adónde vai;
polo mismo que n'o sei,
quixera o poder andar."

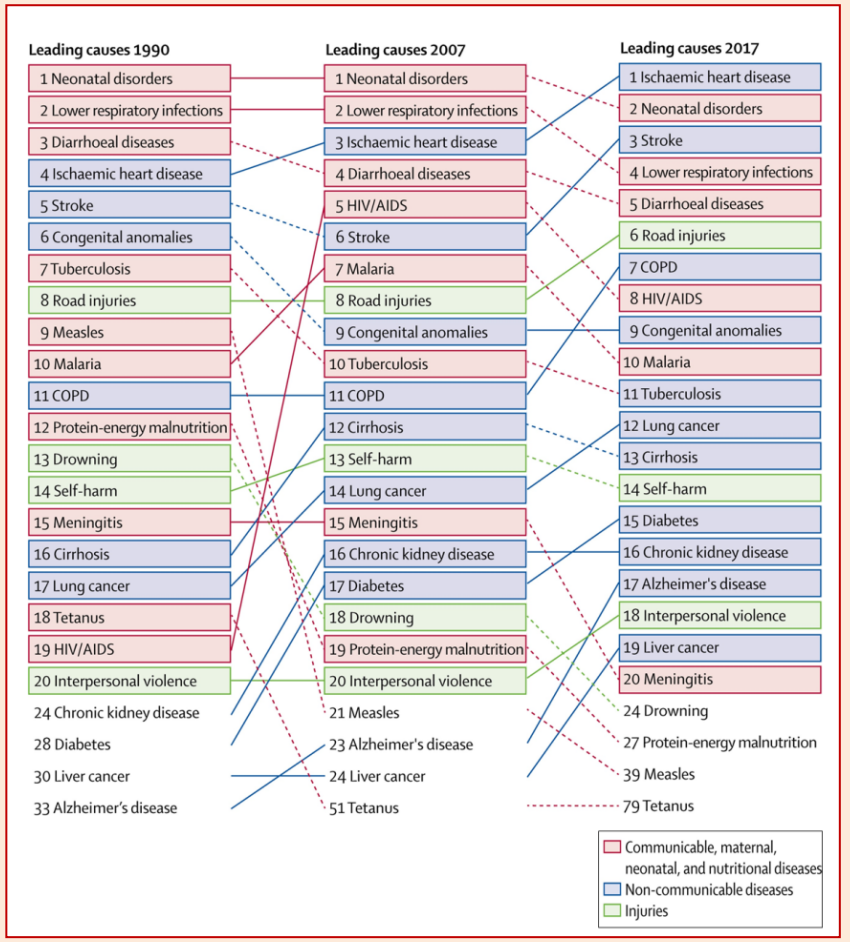
"I can see a road from where I am
and I don't know were it goes;
only because I don't know,
I would like to walk through it."

Rosalía de Castro (1837-1885).

1.1. Cardiovascular diseases

Cardiovascular diseases are the leading cause of mortality worldwide (Figure 1) [1]. They comprise several diseases with different aetiologies that affect the heart or blood vessels. The main clinical manifestations of these diseases are coronary heart disease, stroke or cerebrovascular disease, and peripheral vascular disease. Coronary heart disease (also known as ischemic heart disease or coronary artery disease) is currently the world's biggest health threat, and the main pathological mechanism responsible for all of these diseases is a progressive inflammatory process called atherosclerosis.

Figure 1. Leading causes of global mortality for 1990, 2007 and 2017



Mortality is measured as Years of Life Lost, explained in Box 1.
Image modified from [1].

a) Overview of the burden of cardiovascular diseases

Cardiovascular diseases (CVD), which are categorised as non-communicable diseases (NCD), are complex diseases that result from an interplay between genetic, physiological, environmental, and lifestyle factors. While some of these factors have been explored by many studies in different NCD contexts, the precise underlying biological mechanisms that explain their impact on health are still being unravelled. There are also thought to be some unknown factors that remain to be identified.

Despite medical advances, CVDs have led the global ranking of causes of death every year for the last decade. Currently, two of the clinical manifestations of CVD, coronary heart disease and stroke, are among the top three leading causes of mortality worldwide (Figure 1). Other measures of the burden of disease (see Box 1) also highlight these diseases as the current biggest threat for global health [1–3].

Box 1. Epidemiological indicators used to describe the burden of diseases

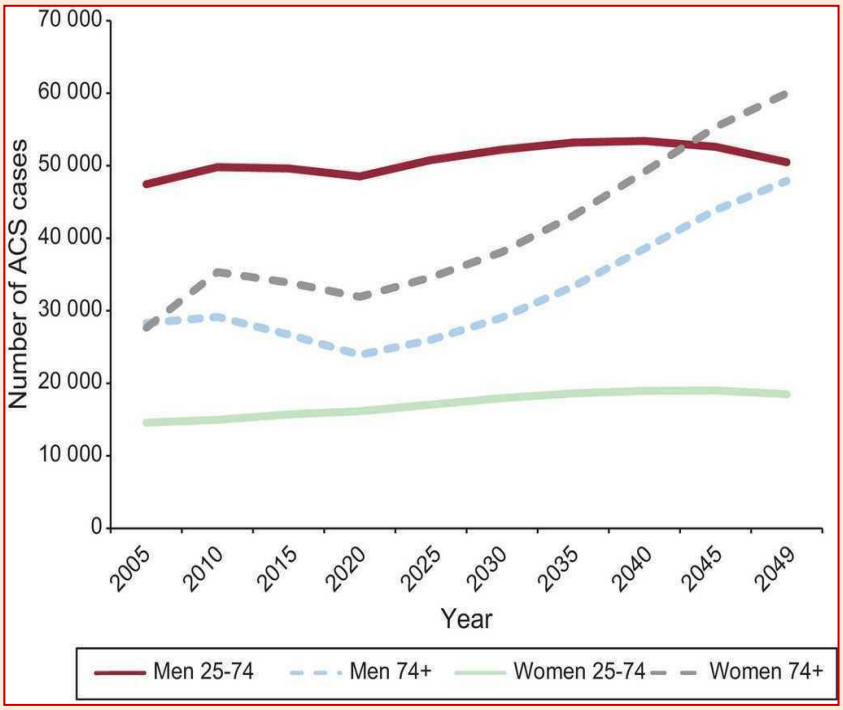
- **YLLs – Years of Life Lost**, a metric of the burden of a disease based on premature mortality
- **YLDs – Years Lost due to Disability**, a metric of the burden of a disease based on the number of years living with a health condition or its consequences
- **DALYs – Disability-Adjusted Life Years**, a metric of the overall burden of a disease considering both YLLs and YLDs

The effects of these diseases on health have driven efforts in cardiovascular epidemiological research and prevention strategies since the mid-20th century. This research is important not only for monitoring the trends, distribution, and frequency of CVDs, but also for deciphering the causes and mechanisms that trigger them. In fact, while CVD had become the main cause of death in the USA of the 1940s, as recently reviewed, "prevention and treatment were so poorly understood that most Americans accepted early death from heart disease as unavoidable" [4]. Among those Americans was US president Roosevelt, who died of stroke due to untreated hypertension and after suffering several cardiovascular events. Closely related to that

event, the Framingham Heart Study (FHS) started in that industrial town in the state of Massachusetts in 1948 [5]. The FHS was the first epidemiological study of its kind. It identified the first factors found to increase cardiovascular risk (CVR), coined the term "cardiovascular risk factor" (CVRF), proposed the first CVR functions for predicting future events [6,7], and opened the door to preventive medicine and a more conscious lifestyle.

Current CVD statistics are alarming, and are predicted to increase in the coming decades as the population increases and ages (Figure 2) [1,3,8–10]. Similar to the paradigmatic transformation of Western societies in the first decades of the 20th century, we have been going through a dramatic change as a society since the late 20th century. The first half of the past century was characterized by key features for health, such as the development of vaccines and antibiotics, improvements in hygiene and sanitation, and Public Health actions. In

Figure 2. Expected trends of acute coronary syndrome (ACS) cases from 2005 to 2049 by sex and age group in the Spanish population



**ACS is explained 1.1.b.*
Image from [10].

addition to the Declaration of Human Rights[1], this led to an epidemiological transition due to a decrease in infectious diseases and a subsequent increase in life expectancy and NCDs [11]. Currently, our society is facing new challenges related to aging and population growth, our behaviour, and our busy and fast lifestyle [8].

Moreover, despite global trends, there are still marked differences between societies. Cardiovascular mortality is reaching a plateau in many countries with high Socio-Demographic Index (SDI)[2], but is still increasing in low- and middle-SDI countries [1]. Some low- and low/middle- SDI areas are experiencing a double burden of disease, with continued high rates of infectious diseases and undernutrition (Figure 3) [3].

Figure 3. Leading causes of DALYs at lowest levels of development, 2017

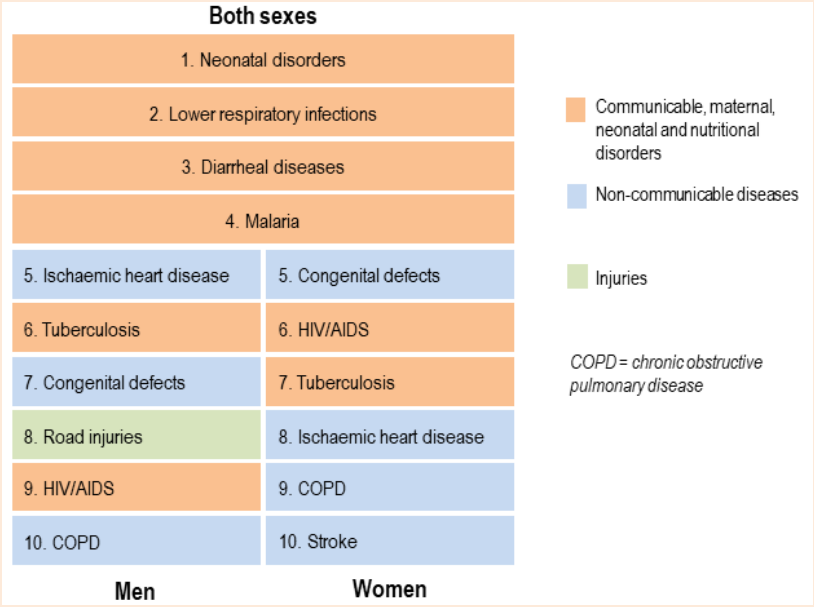


Image created from [3].

[1] The Universal Declaration of Human Rights states that, "Everyone has the right to a standard of living adequate for the health and well-being of himself and of his family, including food, clothing, housing and medical care and necessary social services, and the right to security in the event of unemployment, sickness, disability, widowhood, old age or other lack of livelihood in circumstances beyond his control" (UN General Assembly, 1948.)

[2] SDI, ranging from 0 to 1, measures the level of development of a geographic area based on fertility and average income per person and educational attainment.

b) Coronary Heart Disease

Coronary Heart Disease (CHD) is the leading specific cause of death worldwide [1]. Clinical manifestations of CHD such as myocardial infarction, angina and sudden cardiac death are related to atherosclerosis (further explained in section 1.1.c). CHD mainly occurs from the fifth or sixth decade in men, and from the sixth or seventh decade in women (Figure 4) [1].

Figure 4. Sex differences in global mortality by age, 2017

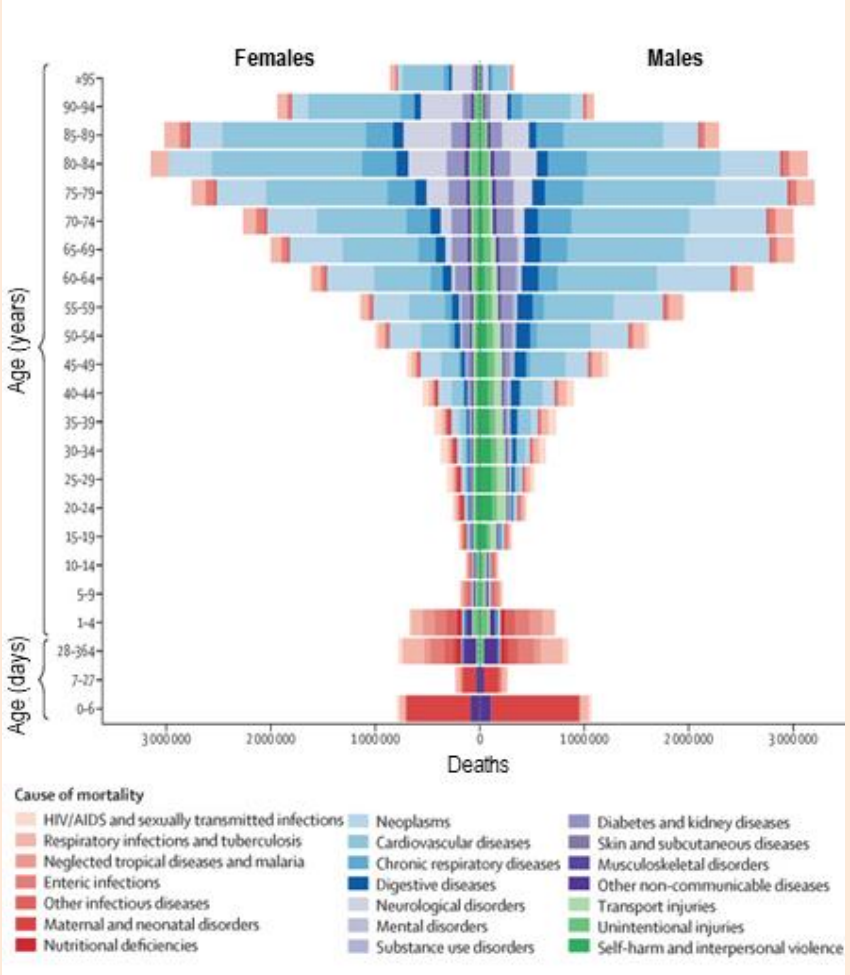


Image modified from [1].

One of the most common presentations of CHD is Myocardial Infarction (MI). The mean worldwide count of YLDs due to MI increased by 13.6% between 2007 and 2017 [2]. MI occurs when myocardial cells die due to occlusion of a coronary artery, impairing blood supply to that area of the heart (known as myocardial ischaemia; Figure 5; [12]). Typical symptoms include oppressive chest pain and discomfort. In women, it more often manifests itself as unusual fatigue, shortness of breath, light-headedness, nausea, or discomfort in the lower chest, upper back, or jaw [13]; some of these symptoms can also be experienced by men. Also more frequently among women, MI can also occur as a silent event (also known as unrecognized MI) [14]. According to the Joint ESC/ACC/AHA/WHF Task Force[3], to diagnose MI, the following features must be observed:

Figure 5. Representative autopsy cases of cardiac rupture as a consequence of fatal AMI

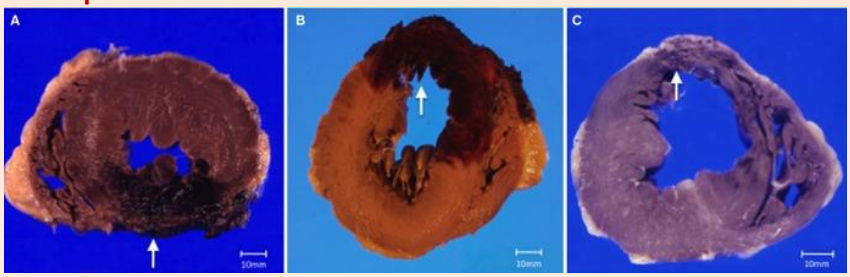


Image adapted from [12].

- Cardiac troponin levels – biomarker of myocardial injury – with at least one value above the 99th percentile, with values remaining abnormal at the first assessment, and after 3 and 6 hours (Figure 6).
- Evidence of the myocardial ischaemia symptoms described above.
- Electrocardiographic changes within 10 min of presentation.
- New loss of viable myocardium or new abnormal regional motion or thickening, as assessed using imaging techniques such as magnetic resonance.

[3] ESC, European Society of Cardiology; ACC, American College of Cardiology; AHA, American Heart Association; WHF, the World Heart Federation.

Figure 6. Scheme of early cardiac troponin kinetics in patients after acute myocardial injury including acute AMI

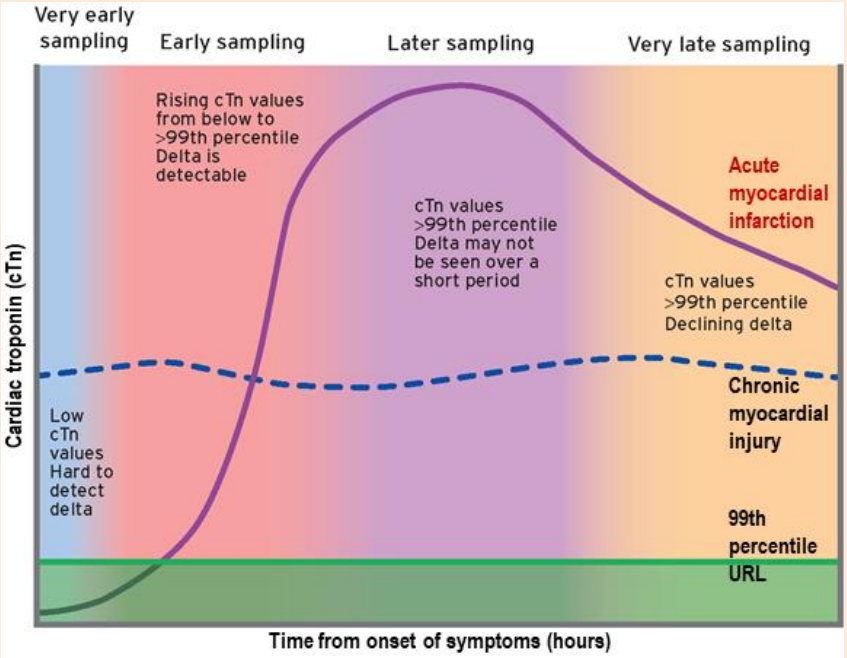


Image modified from [15].

Identification of a thrombus using coronary angiography (or by autopsy) would contribute to a conclusive diagnosis. The Task Force also classifies MI into five types according to differences in its pathology, clinical features and prognosis (see Box 2). Subsequent cardiac management and treatment depends on the type of MI [15].

Acute MI (AMI; MI types I-III) and unstable angina define Acute Coronary Syndrome (ACS). If cardiac troponin levels are normal but symptoms and electrocardiogram changes suggest acute coronary ischemia, AMI is discarded and the acute event suggests unstable angina [15]. This outcome occurs after myocardial ischaemia without myocardial necrosis at rest (thus, cardiac injury biomarkers are not released), and is often the clinical manifestation that precedes an AMI [16]. It differs from stable angina (also known as effort angina), which is due to transitory impairment of blood flow to the heart, usually after physical activity or psychologic stress, and which disappears with rest or medication [17].

Box 2. Classification of Myocardial Infarction

- **Type I**, MI resulting from myocardial ischaemia following plaque disruption with coronary atherothrombosis*
- **Type II**, MI resulting from myocardial ischaemia following unbalanced oxygen demand/supply unrelated to thrombosis*
- **Type III**, fatal MI with symptoms suggestive of ischaemia but without abnormal cardiac troponin*
- **Type IV**, MI resulting from myocardial ischaemia following percutaneous coronary intervention or stent thrombosis
- **Type V**, MI resulting from myocardial ischaemia following cardiac surgery

**Acute events (AMI)*

c) Atherosclerosis

Atherosclerosis is a complex chronic inflammation of the medium and large arteries, resulting in remodelling and narrowing that, in turn, can lead to ischemia or necrosis of certain tissues (e.g. the myocardium) [18]. This silent process starts early in life, even during foetal development [19], with the formation of fatty streaks due to the accumulation and oxidation of deposits of cholesterol and other lipids in the sub-endothelial space of an artery. Other characteristic features include endothelial dysfunction, infiltration of macrophages, formation of foam cells, and activation of T-lymphocytes.

The initial fatty streaks evolve slowly to more advanced lesions characterized by the formation of a fibrous cap that separates the lesion from the lumen. Those with thick regular fibrous cap (stable plaques) are less prone to rupture, while those with areas of thinning of the fibrous cap due to the accumulation of macrophages and pro-inflammatory molecules (unstable or vulnerable plaques) are more prone to rupture. This causes the formation of a thrombus that may occlude the arterial lumen and trigger an acute event [18].

The American Heart Association classically distinguishes six stages of atherosclerosis progression based on the type of histological lesion (Figure 7; [20]) [21].

Figure 7. Progression of an atherosclerotic lesion

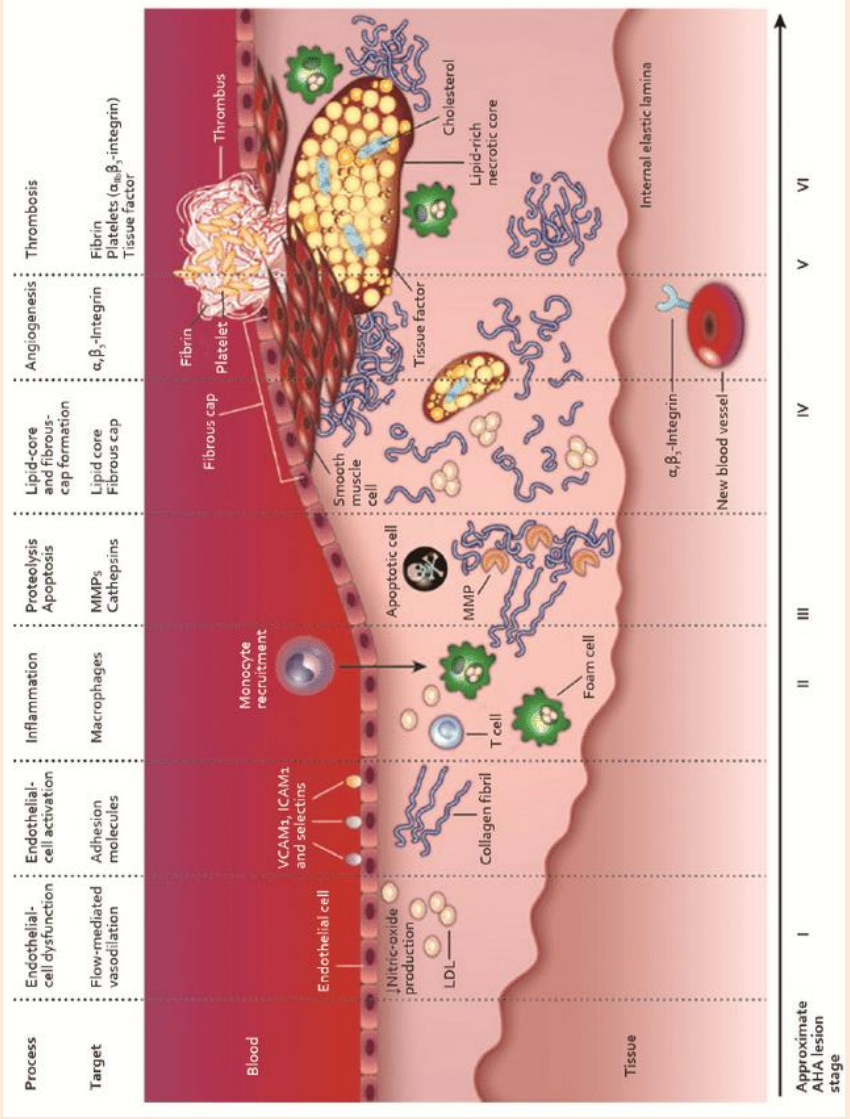


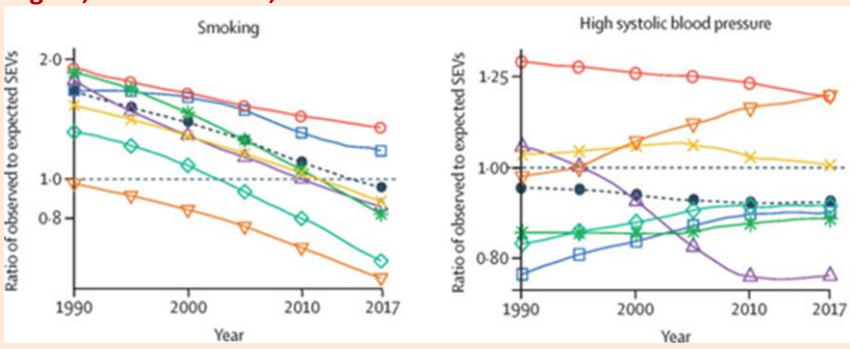
Image from [18].

1.2. Factors that increase cardiovascular risk

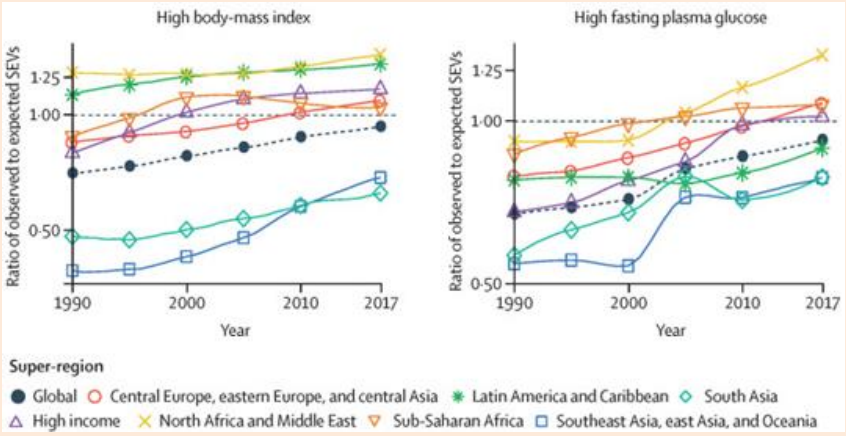
As previously mentioned, the concept of "cardiovascular risk factors" (CVRFs) was introduced by the FHS investigators. These are traits that increase an individual's risk of suffering a cardiovascular event independently of other traits, and can be modifiable or non-modifiable depending on whether we can alter them by changing habits or behaviours, or our environment. They can be hereditary, physiological, environmental and behavioural factors, and they can interact between them.

As Dawber *et al.* mentioned in their editorial about the FHS, "for many years, atherosclerosis and diseases related to it were considered to be the inevitable result of the aging process". Additionally, sex differences in CHD have been observed for a long time [22]. The FHS first identified high lipid levels, blood pressure, smoking, overweight, physical inactivity, and diabetes as factors that increase CVR. They also found that these factors have a cumulative effect on CVR [23]. All of these factors are indeed major contributors to CVR. For example, in the last two decades of the 20th century, the number of CHD deaths in the USA decreased by 44% due to improved management of total cholesterol levels, systolic blood pressure, smoking, and physical inactivity, even despite increased body-mass index and the diabetes [24]. Globally, preventative policies have been partly successful in controlling some CVRFs such as hypertension and smoking. However, the increasing prevalence of some classical and modifiable CVRFs, such as obesity or type 2 diabetes (Figure 8), could lead to increased incidence of CVD in the near future [8,25].

Figure 8. Trends in the ratios between observed and expected (SDI-based) summary measures of exposure according to geographical region, for both sexes, 1990-2017



Continued



**SEVs: Summary Exposure Values. Image adapted from [8].*

a) Aging

The most important factor for CVR is the person's age. As mentioned above, population aging is one of the causes that explains the current and predicted trends in CVD (and NCD in general). As people age, their CVR increases exponentially independently of other CVRFs [26]. In fact, CVD events are less common at younger ages, and 20- to 40-year-olds have low CVR, even if they have high-risk CVRF profiles [27–30]. In the specific case of CHD, atherosclerosis progresses throughout life and its clinical manifestations usually appear from middle-age on [31].

Aging is an extremely complex biological process. The aging-related molecular pathways that are involved in cardiovascular health include mitochondrial dysfunction, deregulated autophagy, angiogenesis impairment, endothelial dysfunction, loss of telomeric DNA, and metabolic cascades triggered by growth factors, FOXO factors and sirtuins [32]. Understanding the fundamental mechanisms that underlie aging could lead to significant improvements in preventive and therapeutic strategies for CVD.

Related to aging and CVR, vascular age has been proposed as a tool for raising awareness about CVR among young individuals [33]. Classically, CVR is expressed as the probability of presenting a

cardiovascular event in the following 10 years. As mentioned, CVR is low in young individuals, even those individuals with high-risk CVRF profile, so it is interesting to express CVR as vascular age, instead of as a percentage risk. This is easily understood by patients in primary care settings, and has a higher impact on the youngest age groups as they can better perceive their CVRF load. Figure 9 shows an example in which a 40-year-old individual has moderate absolute risk (8% at 10-years), but a vascular age of 69 years. Communicating risk to a 40-year-old individual as absolute risk ("you have an 8% of CVR, which corresponds to a moderate risk") has much lower impact than doing so using vascular age ("your chronological age is 40, but your vascular age is 69 years old") [34]. For instance, communicating vascular age was shown to improve the control of the lipid profile [35].

Figure 9. Cardiovascular risk expressed as vascular age

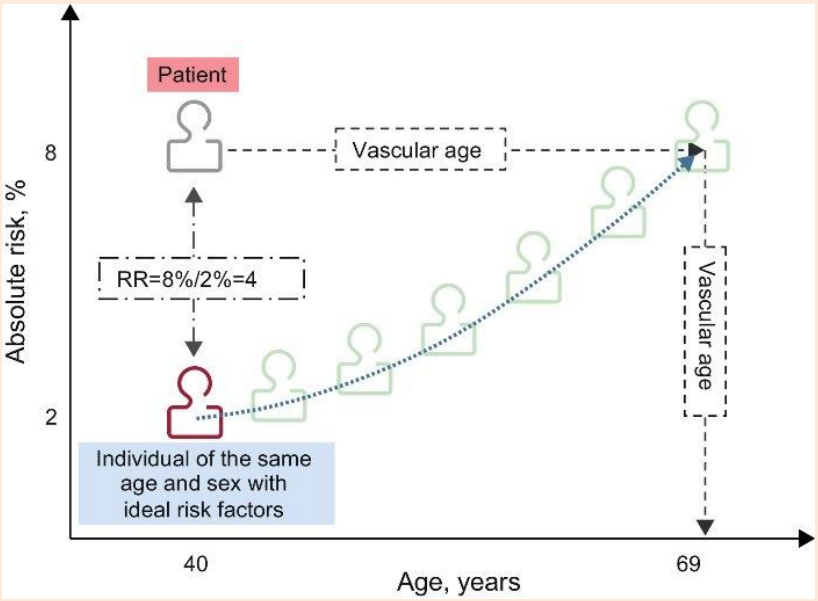


Image from [34].

b) Sex

The second most important non-modifiable CVRF for CVD is the sex of the individual. CVD death is more common among women [1], as a consequence of a higher incidence of stroke, whereas CHD deaths are

more common among men [1]. However, CHD may be underdiagnosed in women because of its less alarming symptomatology. In addition, as mentioned above, women usually have CHD and overall CVD events at older ages than men. Moreover, CVR estimates that consider both age and sex quite accurately differentiate individuals that will have a CVD event from those who will not [26].[4] Finally, another important sex-related issue for CVR is the fact that the effect sizes of other CVRFs differ between women and men. Some of these differences will be explained in further detail below in the sections on CVRFs. The differences between sexes have been attributed to a protective effect of female sex hormones [36].

c) Ethnicity

The FHS included Americans of European ancestry, and, in fact, traditionally epidemiological studies have not equally represented all ethnic backgrounds. However, racial/ethnic groups vary greatly in CVR and in the prevalence or magnitude of some CVRFs. For instance, African Americans have higher risk of CHD death [37] and higher blood pressure and hypertension rates than Americans with a different ancestry [38,39]. Similarly, UK citizens with an South Asian ancestry also present higher CVR than those of European ancestry [40].

d) Smoking

Tobacco smoking is a modifiable CVRF that causes endothelial dysfunction and increased risk of thrombosis [41]. Smoking was shown to increase risk of CVD death, and to have a non-linear dose-response association with relative risk of CHD, with higher numbers of cigarettes smoked leading to higher CHD risk [42]. In Spain, MI incidence and death diminished in an 11% after the introduction of a partial smoking ban in 2006 [43].

[4] This capacity to differentiate individuals who will suffer a CVD event from those who will not is explained in the next section. It is called discriminative capacity, and is measured with the C-statistic index, which ranges between 0.70-0.75 in this specific case.

Differences in CVR are not only observed between smokers and non-smokers, but also between current non-smokers who previously quit and those who have never smoked. Smokers have higher risk of MI and CVD death than former smokers [44,45], and CVR decrease quickly after quitting smoking [46]. However, former smokers who quit smoking more than 10 years previously still have higher CVR than people who never smoked [46]. Also, individuals who began smoking at or before 12 years of age have an inverse and linear association between aged of onset of smoking and CVD outcomes [47].

Smoking also leads to higher CVR among individuals exposed to second hand smoke, who are reported to have 25-30% higher CHD risk than non-exposed individuals [48]. In addition, women who smoke during pregnancy expose their children to this CVRF, resulting in higher CVR in adulthood than among individuals who had not been exposed to maternal smoking as a foetus [49].

Smoking varies between women and men, with women traditionally starting smoking later and smoking fewer cigarettes than men. However, the age of onset of smoking and daily doses are currently similar in both sexes [36]. In addition to gender differences in smoking habit, smoking-related CHD risk is higher in women [50], at least when comparing individuals who smoke >20 cigarettes per day (known as "heavy smokers").

Since smoking is a modifiable CVRF, public health and policy actions contribute to social awareness of the risks of smoking and its prevention. In the Spanish example mentioned above, the partial smoking ban not only influenced smokers to smoke less [51], but also reduced the prevalence of second hand smokers [52].

e) Dyslipidaemia and abnormal levels of blood lipids

Circulating levels of triglycerides (TG) and cholesterol are a modifiable CVRF that is closely related to atherosclerosis. Both lipids are required by all cells in the body, and are therefore transported throughout the body in the bloodstream, coupled to apolipoproteins because of their apolar nature. The joint lipid-apolipoprotein macromolecules are called lipoprotein particles, which are grouped in four classes depending on their molecular weight: very low density lipoproteins

(VLDL), intermediate density lipoproteins (IDL), low density lipoproteins (LDL) and high density lipoproteins (HDL). Most common dyslipidaemias consist on an elevation of TG levels (hypertriglyceridaemia) or an alteration on cholesterol levels (hypercholesterolemia and HDL-related alterations).

• Hypertriglyceridaemia

Hypertriglyceridaemia is defined as levels of circulating TG higher than 150 mg/dL [53]. Serum TGs are mostly carried by VLDL particles, and these TG-rich lipoproteins can enter the arterial wall and boost atherosclerosis [54]. A meta-analysis of some of the first studies on TG as an independent CVRF reported relative CHD risks of 1.32 and 1.76 per mmol/l increase in serum TG levels in men and women, respectively [55]. Mendelian Randomisation[5] studies subsequently showed that this is a directly causal association [55–57].

• Hypercholesterolemia

This health condition is characterized by high plasma levels of total cholesterol and/or cholesterol carried in LDL particles (LDL-C) or non-carried in HDL particles (HDL-C). In some cases it has a genetic basis and appears during childhood, known as familial hypercholesterolemia. However, most cases can be treated through lifestyle changes.

▪ High total cholesterol levels

The total amount of cholesterol in the blood (in any lipoprotein) is directly associated with higher CHD risk. The reference values for normal total cholesterol levels are <200 or <240 mg/dL according to different guidelines. Measured in men early in their adulthood, it has been shown to be strongly related to incident CHD and CVD and to CVD death later in life, independently of other CVRFs [58].

[5] Mendelian randomisation studies are addressed in 1.3.b.

- **High LDL-C levels**

LDL-C is transported to the cells. Healthy levels of LDL-C are <100 mg/dL [59]. High levels of circulating LDL-C are linearly and causally associated with higher risk of CHD. Randomized controlled trials of therapies for reducing lipid concentrations, namely statins, ezetimibe, and PCSK9 monoclonal antibodies, have been shown to reduce CHD risk [60], supporting their causal role in CHD risk. Mendelian Randomisation studies also support this causal association [57].

• Low HDL-C levels and impaired HDL functionality

Cholesterol that exceeds a cell's needs is transported away from the cell in HDL particles and delivered to the liver for metabolization or excretion, a process known as reverse cholesterol transport [61]. Consistent with this function, HDL-C is inversely associated with CHD risk [62], and also has anti-atherosclerotic properties [63]. Healthy levels of HDL-C are >50 and >40 mg/dL in women and men, respectively, and low HDL-C levels are therefore a CVRF. However, an experimental studies using drugs to increase HDL-C did not show a corresponding decrease in CHD risk [64], and, along with Mendelian Randomisation studies, this result questions the causal association between HDL and CHD [57,65]. These findings suggest that it is not useful to examine HDL-C levels, so current research focuses on qualitative HDL traits to explain its anti-atherogenic role, such as cholesterol efflux capacity or HDL's antioxidant and anti-inflammatory properties [63].

Lipid profile varies between sexes: while women usually have a healthier lipid profile at younger ages than men, after menopause lipid levels increase to higher levels than those observed in men [36].

While some genetic characteristics can module lipid profile, it is a modifiable CVRF. A healthy lifestyle, and specifically a healthy diet, can improve it to a more favourable state. For instance, the Mediterranean diet, which is especially enriched with virgin olive oil, improves the anti-atherogenic functions of HDL [66].

f) Hypertension and high blood pressure

Hypertension is a chronic, symptomless condition that is characterized by persistent high blood pressure (BP). More precisely, according to European guidelines, it is defined as a resting systolic blood pressure (SBP) of ≥ 140 mmHg, or resting diastolic blood pressure (DBP) of ≥90 mmHg [30]. Both SBP and DBP are linearly related to CVR. Antihypertensive treatment lowers risk of stroke, MI, and heart failure by 35-40%, 20-25%, and >50%, respectively [67]. Thus, the higher the BP, the higher the risk, independently of what BP cutpoints (thresholds) are used to define hypertension. Current American guidelines have changed these cutpoints to 130 mmHg for SBP and 80 mmHg for DBP [68].

Hypertension differs between sexes, with higher incidence in men until the sixth decade of life [69]. Also, a large fraction of hypertensive individuals are unaware of it, and the disease is insufficiently controlled in many who are aware of their condition [70]. Thus, improving this CVRF depends on raising social awareness and controlling it better.

g) Diabetes and high glucose levels

Diabetes is a complex metabolic disease characterized by high levels of fasting serum glucose (known as hyperglycaemia). Specifically, diabetes is defined as glucose levels of ≥126 mg/dL. Type 1 diabetes results from a loss of pancreatic β-cells and impaired insulin production, and is usually diagnosed at young age. Conversely, type 2 diabetes (T2D) is a consequence of insulin resistance, i.e. impaired cell response to insulin, and typically appears in middle age. A third type of diabetes is that which occurs in pregnant women who develop hyperglycaemia without a previous history of diabetes, and is called gestational diabetes.

Diabetes has been shown to increase CHD risk by 2-3 fold [67], mainly based on studies of T2D patients [71,72]. However, for the first time, the 2017 Global Burden of Disease study disaggregated diabetes to better estimate DALYs [3]. Thus, from now on in this thesis, when describing diabetes, I will be referring to T2D.

Diabetes is closely associated with other CVRFs. On the one hand, there is a clear sex differences in CVR among diabetes patients, with diabetic women having higher relative risk of CHD and stroke than diabetic men (44% and 27%, respectively) [50,73]. On the other hand, diabetes is also associated with hypertension, hypertriglyceridemia, low HDL-C levels, and obesity, as all of these are related to insulin resistance. The metabolic syndrome is defined according to a combination of these CVRFs (further explained in 1.2.h).

Despite its complexity and accounting for one of the highest increases in DALYs between 2007 and 2017, diabetes is a modifiable CVRF. Lifestyle interventions that achieve a small amount of weight loss have been shown to decrease diabetes incidence by 58% after 3 year, and by 34% after 10 years [74–76].

h) Overweight and obesity

Obesity is an abnormal accumulation of body fat related to several health problems, and is considered a complex medical condition of epidemic magnitude [3]. It causes altered lipid and glucose metabolism, and impairs respiratory and cardiovascular function and structure. The most common measure used to define obesity is body mass index (BMI), which is computed from weight and height (BMI = weight [kg] / height^2 [m]). BMI categories are underweight (<18.5 kg/m^2), normal weight (18.5-24.9 kg/m^2), overweight (25.0-29.9 kg/m^2), and obesity (≥30.0 kg/m^2). Obesity is further classified in three classes based on BMI [77]:

- class I obesity: 30.0-34.9 kg/m^2,
- class II obesity: 35.0-39.9 kg/m^2, and
- class III obesity: ≥40 kg/m^2.

A more specific indicator of obesity is abdominal obesity, based on waist circumference, which reflects visceral fat accumulation rather than subcutaneous deposits. Visceral fat is linked to inflammation that is in turn linked to metabolic diseases and CVD. While there is some controversy regarding what cut-off values are appropriate to define this indicator, currently individuals with waist circumference >88 and 102 cm in women and men, respectively, are considered to have abdominal obesity [78].

The FHS found that obesity is associated with CVD risk in a dose-dependent manner: two years with obesity increases risk of CVD-related mortality by 7% [79]. In addition to the link between obesity and the other CVRFs mentioned, there are differences in CVR between obese women and men, and obesity accounts for 6% and 5% of CHD deaths in women and men, respectively. Obesity-related CVR also depends on age, with overweight and obese children having higher CVR in adulthood.

The close correlation between obesity, diabetes, lipid profile, and BP led to the definition of the metabolic syndrome, which is diagnosed in individuals who have a combination of at least three of these modifiable CVRFs, as follows [80,81]:

1. Fasting plasma glucose levels >100 mg/dL.
2. Hypertriglyceridemia (TG levels >150 mg/dL)
3. Low HDL-C levels (<50 and 40mg/dL in women and men, respectively).
4. Hypertension (SBP >130 mmHg, DBP >85 mmHg).
5. Abdominal obesity.

As these traits are each individually related to CVR, metabolic syndrome is also unsurprisingly associated with higher risk [82]. However, it is still unclear whether the definition of metabolic syndrome as defined above provides more information for estimating CHD risk than simply the sum of the individual traits [83].

Obesity is considered a major public health problem due to its increasing prevalence [84], which then leads to a higher prevalence CVD, diabetes, hypertension, and other chronic diseases such as cancer [3]. Thus, it is urgent to implement public health and policy actions to counteract this epidemic [25].

i) Physical inactivity and sedentary behaviour

Physical activity (PA) has multiple benefits on health throughout life [85,86]. Current international recommendations for physical activity in adults are based on aerobic PA, and recommend at least 150 minutes of moderate-intensity PA or 75 minutes of vigorous-intensity PA per

week [87]. The precise mechanisms triggered by PA are still not well understood, although it is known to cause anti-atherogenic adaptations in vascular function and structure, and to decrease chronic inflammation, independently of other CVRFs [86].[6]

Insufficient PA is a CVRF as well as a risk factor for other health problems related to CVD, such as diabetes or obesity. The FHS reported that more sedentary participants were more prone to fatal MI events [88]. Conversely, risk of CHD death was found to be inversely associated with categories of occupational energy requirement among longshoremen [89]. More recently, meta-analyses using continuous PA metrics reported that changing from inactivity to meeting the current PA recommendations lead to a 23%, 17% and 14% decrease in risk of CVD death, CVD, and CHD, respectively [90,91]. Even individuals who did not reach the recommended levels had lower CHD risk than physically inactive individuals, and at higher doses of PA, the benefits for cardiovascular health were higher [91]. Also, the exercise regimes used in cardiac rehabilitation of ACS have also been reported to reduce risk of CHD death [92,93].

Traditionally, studies have focused on the benefits of PA compared its absent or insufficient activity. However, more recent research has addressed sedentary behaviour, assessed as time spent sitting, and found that it has an independent effect on CVR, beyond that of insufficient PA. Sedentary behaviour is associated with higher risk of CHD, CVD and CVD mortality [94–98].

Humans are genetically adapted to a more active lifestyle than we have currently, when PA and sedentary behaviour have achieved their lowest and higher levels in human history, respectively. PA is decreasing and sedentary behaviour increasing even more in younger generations, and this trend raises concerns about its impact on general health. However, PA is an underused preventive strategy [99], and sedentary behaviour is not adequately discouraged. Thus, policy makers and health professionals should prioritize actions that target more inactive individuals, enhance their adherence to PA, and advise them on reducing time spent sitting. Population-level approaches for each age-group should also be pursued [100,101]. However,

[6] PA will be addressed in one of the objectives of this thesis.

recommendations should consider that excessive PA may increase risk of osteoarthritis [102], sudden cardiac death and AMI in susceptible persons [103].

j) Other cardiovascular risk factors

Other lifestyle and environmental factors have been reported to modulate CVR, including: diet [104,105], changes in gut microbiota [106], alcohol consumption [107], socioeconomic status [108], psychological stress [109], lack of sleep [110], and air pollution [111]. In addition, other diseases are associated with higher CVR, such as chronic kidney disease [112]. Another non-modifiable classical CVRF is family history of CVD. While high CVR can be a consequence of a family's cultural factors (i.e. common habits and environment), this risk is also partly and independently explained by a hereditary component. The Genetics of CVD is further described in a later section.

1.3. Prediction of the individual cardiovascular risk

Prevention of CVD in the general population requires both population-based and individual preventive strategies. The first approach is based on the modifiable nature of some CVRFs, and as already discussed, is achieved by promoting healthy or favourable lifestyles via community actions and policies. Conversely, individual strategies focus on assessing individuals' CVR and implementing individual prevention strategies, the intensity of which varies according to the estimated CVR in order to prevent the occurrence of a first event, which may be fatal [34].

a) Cardiovascular risk functions

Cardiovascular risk functions are the recommended tool for estimating individual cardiovascular risk. In cooperation, the clinician and the individual implement preventive measures that are more or less intensive depending on the estimated risk [30]. Technological advances allow mass population screening, but currently there is no evidence to support this as a cost-effective prevention strategy [113]. Therefore, screening based on risk estimation is limited to certain individuals, in order to concentrate the greater intervention effort on individuals with the highest risk. Population screening is done in individuals who contact the health system for any reason (opportunistic screening), and also in those with a family history of CHD events or hypercholesterolemia, or who have clinical CVRFs (high risk screening). In both types, the most common approach is to use CVR functions and charts, which generally estimate the individual's CVR during the following 10 years [113,114].

CVR functions are equations that estimate the probability of having a CHD or CVD event within a certain period of time [34]. The most common CVR functions used in clinical and epidemiological settings are based on classical CVRFs, as they are strongly and independently related to CHD. At this time, no other CVR-associated factors improve the predictive capacity of standard CVR functions. The resulting probability is the absolute risk, which can be further categorized into low, intermediate, high, or very high risk [34].

While the absolute risk provided by the CVR function has a straightforward interpretation, in that it represents a percentage, the clinician-patient dialogue is considerably more complex. First, the clinician cannot anticipate whether their patient will be one of those who actually will have a cardiovascular event, out of 100 individuals with the same load of CVRFs. In addition, neither this probability nor the assigned risk category are always properly understood by the patient, such that they immediately and faithfully modify their habits. This is especially observed with young patients, who usually have low or intermediate risk, even with an unfavourable burden of CVRFs, as mentioned in the previous section. In this context, it is highly recommended to transmit their predicted risk as a relative risk, i.e. that their risk is x times higher than if they had an optimal CVRF profile, or as vascular age (Figure 9) [34].

CVR functions developed for one specific population can generally be adapted for another. To do so, the CVR function needs to be adequately calibrated, i.e. the CVR estimation must reflect the corresponding epidemiological situation in terms of the incidence of CHD/CVD and of prevalence of CVRFs. Also, where the epidemiology of the disease and/or CVRFs change in the original population, the risk function must be recalibrated [34]. CVR functions must also be validated, which requires a prospective cohort. Validity is assessed using two components: i) accuracy, in which the number of predicted events is compared to that observed during follow-up using a goodness-of-fit test that informs about the calibration (expected=observed); and reliability, which refers to the function's capacity to discriminate individuals who will suffer an event from those who will not. The discriminative capacity is calculated using metrics such as the c-statistic [115].[7]

- **The Framingham risk function and its adaptations**

As mentioned above, the first CVR function was developed by the FHS [7], and was updated at the end of the 1990s. [114]. During the past two decades, this current version has been adapted and calibrated for other populations, such as the Spanish one [116–118]. All of these

[7] Both calibration and discrimination metrics will be further described later.

functions estimate the risk of presenting fatal or nonfatal CHD within 10 years in individuals aged 35-74 years, and are based on:

- the incidence of CHD in the population (1-S),
- the individual's CVRFs load,
- the population means of the CVRFs load ($CVRF_p$), and
- the effect of each CVRF on CHD risk (β_{CVRF}).

In population-adapted functions, the incidence of CHD and the prevalence of the CVRFs in the original function are substituted with their corresponding values in the target population. The effects of each CVRF are transferred directly [117].

• Cardiovascular risk functions in Spain

Many high risk patients are not correctly managed, while lower risk individuals receive drug treatment [34]. This observation highlights the incorrect or insufficient use of primary prevention strategies in the Spanish clinical settings. In fact, only 38% of clinicians admit that they calculate CV risk in >80% of patients with at least one CVRF [119]. It seems likely that this situation would improve if there was consensus on which CVR function to use [120]. However, as stated by the Spanish Interdisciplinary Committee on Cardiovascular Prevention, "independent of the function used, the important issue is to estimate CVR as a strategy for cardiovascular prevention in clinical practice, and to apply current guidelines according to clinical criteria and the patient's preferences". Three main functions are used in Spain (in order of increasing use): the SCORE function (and charts), the original Framingham CVR function, and the REGICOR function (and charts) [120]. Another function adapted to the Spanish population is the FRESCO function, which is also listed in the Spanish Guidelines mentioned above [27,121].

▪ The SCORE function and charts

SCORE is a European CVR function developed by the Systematic COronary Risk Evaluation (SCORE) Project [122]. Its use is recommended by the most recent European guidelines on

cardiovascular disease prevention in clinical practice [30], and by the Spanish Interdisciplinary Committee on Cardiovascular Prevention [121]. SCORE estimates 10-year risk of fatal CVD in 35- to 64-year-olds. In fact, there are four SCORE functions, calculated for regions of Europe with high and low CVR, and using two models of estimation, one based on total cholesterol, and another on the ratio of total cholesterol to HDL-C. To make them more user-friendly, risk estimates based on SCORE are displayed as risk charts including 4 risk categories: low (<1%), moderate (1%-4.9%), high (5%-9.9%) and very high (≥10%).

While SCORE has been calibrated for the Spanish population [28], a recent assessment of its validity (reliability and accuracy) by the FRESCO study (see below) found that none of the three SCORE functions – the original low-risk SCORE function with and without HDL-C and its calibrated version for the Spanish population – could accurately predict CVD mortality (it was overestimated), indicating that this function needs to be recalibrated for the Spanish population. In the FRESCO study, the authors note that "it is thus reasonable to also promote the use of the functions validated in Spain: Framingham-Wilson calibrated by REGICOR and by FRESCO" [123].

- **The REGICOR function and charts**

The REGICOR function is an adaptation and calibration of the Framingham CVR function for the population of the province of Girona (Spain) using data from the REgistre GIroní del COR (REGICOR) Cohort [116], in which it has been validated for estimated 5- [124] and 10-year [125] CHD risk. It was shown to better classify high-CVR patients treated with statins than the original Framingham and SCORE functions [126].

The original Framingham risk categories were adapted to low (<5%), moderate (5%-9.9%), high (10%-14.9%), and very high (≥15%), and the REGICOR estimates are also displayed as risk charts [125,127]. This validated adaptation of the Framingham CVR function is used in three Spanish autonomous communities [120].

- **The FRESCO function and charts**

  More recently, a set of CVR functions were developed and validated for the Spanish population in the 35-79 years age range. FRESCO means "Spanish risk function of coronary and other cardiovascular events", as it consists of a set of risk functions that predict CHD, stroke, and global CVD within 10 years. They use implemented CVRF measurements and can be automatically calculated by electronic medical records systems. In individuals aged ≤74 years old, the FRESCO CHD function discriminates as well as the Framingham-REGICOR function, but did not overestimate CHD risk in the validation cohort, as Framingham-REGICOR tended to [27].

Despite the high value of CVR functions in clinical practice, they all have one important limitation, their low sensitivity[8]. In fact, many CHD or CVD events occur in the intermediate risk category, which accounts for a large percentage of the population. This low sensitivity suggests that these functions have poor predictive capacity in this population subset. In addition, CVR functions only consider some CVRFs independently of either the length of the exposure to those factors or the prescribed treatments. Therefore, it is essential to find new factors that improve the predictive capacity of CVR functions, especially those that can reclassify individuals who are currently classified in the intermediate-risk category [34].

b) Cardiovascular biomarkers

Biomarkers are measurable indicators of a biological condition or a pharmacologic response to a therapeutic intervention. Depending on what they indicate, they can be classified as predictive, diagnostic, prognostic, or therapeutic biomarkers.

As mentioned above, risk functions are useful instruments but present a low sensitivity, so there is a need to explore potential new

[8] Sensitivity (true positive rate) measures the proportion of true positives that are correctly identified as such; here, the proportion of individuals in the correct risk category.

biomarkers to improve their accuracy and individual risk stratification. For this reason, I will focus mainly on predictive biomarkers. The evaluation of a new predictive biomarker consists of sequential steps, analogous to the developmental phases of a new drug (see Box 3) [128,129]. Proof of concept is usually based on the association between the biomarker and the outcome in case-control studies. Prospective validation replicates the association in prospective cohort studies. To evaluate the biomarker's incremental value for the predictive capacity of the risk function, several measures of the function's performance must be considered (see Box 4) [130].

Box 3. Criteria to evaluate novel biomarkers

- **Proof of concept:** different levels found in individuals with and without a certain outcome
- **Prospective validation:** association with developing the outcome
- **Incremental value:** predictive information beyond that provided by existing biomarkers
- **Clinical usefulness:** improvement of current clinical action (e.g. therapy)
- **Clinical outcomes:** improvement of outcomes to a healthier state, preferably in a randomized trial
- **Cost-effectiveness:** justification of additional costs
- **Ease of use:** allowance of its widespread application
- **Methodological consensus:** standardized measurements to facilitate comparison
- **Reference/cut-off values:** definition of reference values to facilitate interpretation

Adapted from [128,129]

Traditionally, research into cardiovascular biomarkers has focused on enzymes and serum biomolecules, such as high-sensitivity C-Reactive Protein, troponin or brain natriuretic peptide. In recent years, several different types of predictive biomarkers have been proposed, but it remains to be seen whether including them in a CVR function will improve its CVD predictive capacity [131–134].

Box 4. Measures of a functions' performance

- **Calibration:** how the risk estimation matches the actual incidence
  - *Calibration graph and Calibration slope*
  - *Hosmer-Lemeshow test*
- **Discrimination:** whether the function differentiates between those who will suffer an event and those who will not
  - *C-statistic (or Area Under the ROC* curve; AUC)*
  - *Discrimination slope*
- **Reclassification:** comparison of risk classifications from the function with and without the biomarker
  - *Reclassification table and reclassification calibration*
  - *Net reclassification Index – NRI*
  - *Integrated Discrimination Index –IDI*

**Receiver Operating Characteristic.*
Adapted from [130]

• Biomarkers of subclinical atherosclerosis

As atherosclerosis is a progressive and silent disease, biomarkers that assess its state of progression using non-invasive imaging techniques may partly overcome the limitation of the long delay between the start of atherosclerosis and its clinical manifestation decades later [135]. For instance, the Multi-Ethnic Study of Atherosclerosis (MESA) Study assessed the predictive value of subclinical atherosclerosis and other risk markers in a multi-ethnic population [136]. Thus, subclinical atherosclerosis biomarkers have been gaining great interest during the past decade, thanks to promising advanced imaging technologies such as magnetic resonance imaging [124], positron emission tomography, and computed tomography. Several measurements are used as a proxy of subclinical phenotypes of the disease, such as intracoronary calcium, the carotid intima-media thickness, atherosclerotic plaques, and the arterial stiffness [115,137]. Some markers are also associated with inflammation and stress biomarkers [138–140], or with traits such as vitamin D deficiency or telomere length [141,142].

- **Coronary artery calcium**

Accumulation of calcium in the internal walls of the coronary artery is a direct sign of atherosclerosis, reflects the CVRF exposure, and it is strongly related to incident CHD. This trait improved risk prediction of the Framingham function, especially in individuals with intermediate risk [137,143]. However, it can only be detected by a computed tomography scan, which is a complex and expensive technique that, moreover, exposes patients to high levels of radiation [144].

- **Carotid intima-media thickness**

Measuring this biomarker – by ultrasound imaging – is more affordable and less aggressive for patients than measuring intracoronary calcium [144]. In addition to being associated with incident MI and stroke [145], it improves CVR reclassification by the Framingham risk function when measured as the maximum value in the internal carotid artery [146]. However, measured thicknesses differ between observers, ecographic devices, scan protocols, softwares, and states of plaque progression [147]. Also, slowed progression of this indicator did not reflect a reduction in CVD events [148]. Recent studies suggest that quantitative measures of plaques (number, thickness, area, and 3-dimensional volume) increase the sensitivity of CVR functions more than the mere presence or absence of plaques, or the thickness of the intima-media [149].

- **Arterial stiffness**[9]

Arterial stiffening is a marker of endothelial dysfunction, and reflects progression in the formation of atheroma, as well as the presence of excessive collagen production and deposition in the arterial wall, leading to atherosclerosis progression. Measures that reflect arterial stiffness include the arterial distensibility coefficient, pulse wave velocity, and the ankle-brachial index, all of which are simple, reproducible, and affordable. They have

[9] Arterial stiffness will be addressed in one of the objectives of this thesis.

been shown to be associated with CVRF profile and with incident CHD events [150–154], but their predictive value for CVR has not yet been fully explored [144].

• Genetic biomarkers

As mentioned above, atherosclerosis and CVD outcomes are complex diseases with a genetic component. Scientific and technological advances have led to the discovery of specific genetic markers associated with cardiovascular health.

▪ Identification of genetic marks of disease throughout the genome

The confluence of human genetics and epidemiology has made it possible to investigate variants in the DNA sequence that are associated with a disease. The most common analysed genetic variants are Single-Nucleotide Polymorphisms (SNPs). A genetic polymorphism is "the occurrence in the same population of two or more alleles at one locus[10], each with appreciable frequency", as defined by Cavalli-Sforza and Bodmer [155]. This "appreciable frequency" was arbitrarily set to at least 1% of the population to distinguish polymorphisms from rare variants[11] [156]. SNPs are specific base-pair positions in the genome in which a certain nucleotide is present for most individuals of a population (major allele), but where at least 1% of individuals have another nucleotide (minor allele, which generally is the risk allele) [157]. In fact, several alleles can be found at one SNP.

Before the year 2000, the genetic variants related to a disease were generally identified by linkage analysis in families with at least one index individual presenting the disease and non-affected relatives, and by candidate-gene association analyses, typically in case-control studies. However, the genetic coverage of these studies was low and most of the common genetic variability was not

[10] Locus (plural loci) is a fixed position on a chromosome. A genetic variant at a given locus is an allele.
[11] Currently, that frequency cutoff is controversial [156,377].

assessed. Since the publication of the full human genome sequence with more than 20,000 genes[12] [158–160], millions of SNPs have been identified throughout the genome by successive efforts, including the International HapMap Project [161] and the 1000 Genomes Project [162]. These projects were the foundation of population-based, hypothesis-free approaches to studying the genetic epidemiology of complex diseases.

- **Genome-wide association studies**

A Genome-Wide Association Study (GWAS) is a hypothesis-free association study between SNPs distributed throughout the genome and a trait of interest. GWAS have allowed researchers to identify genetic variants associated with complex diseases in populations of hundreds or thousands of individuals [163]. An online database summarising information from published GWASs is available at (the GWAS catalog: https://www.ebi.ac.uk/gwas/).

Two main factors converged to drive the GWAS era: the completion of the HapMap Project (which facilitated the linkage disequilibrium-based design of SNP genotyping arrays covering most of the known common variability in the human genome), and the development of technology to develop and analyse these SNP genotyping arrays [164]. The first large-scale association study involving thousands of randomly selected SNPs distributed across the genome was gene-based and was published in 2002 for MI [165]. However, the first collaborative and large GWAS using an array with good coverage of the genome was published five years later for seven complex diseases (including CHD) [166], and the first three GWAS for CHD were also published in 2007. Since then, thousands of GWAS of different phenotypes have been performed. As of September 2018, the free online Catalog of published GWAS[13] contains 5,687 publications and 3,567 associations (https://www.ebi.ac.uk/gwas/diagram) [167].

[12] Eighteen years after the first drafts of the human genome, we still have not found all our genes, and the number varies with the redefinition of "gene" [378].

[13] A high-quality curated collection developed by the US National Human Genome Research Institute and the European Bioinformatics Institute from the European Molecular Biology Laboratory.

GWAS are performed in two steps using two independent population samples. In the discovery phase, SNPs are identified as being potentially associated with the disease, and in the replication phase these SNPs are confirmed or rejected as being associated with the disease [128]. Testing many associations in the same study increases the chance of false positive findings, so GWAS requires multiple correction approaches, large samples of well-phenotyped individuals, and consequently, large international collaborations. Genotype imputation has enabled the reconstruction of variants that are not directly assessed by the array by comparing each sample to a reference panel of sequenced genomes, increasing the statistical power of the analysis, and allowing researchers to combine results across studies [168,169]. This collaborative work between different research groups has highlighted the need for data sharing in public databases to speed up research on complex diseases [169].

- **Next generation sequencing studies**

Although GWAS are hypothesis-free approaches, they do not assess the entire genome, just the SNPs available in commercial arrays. Moreover, they rely on genotype imputation, which requires a fully sequenced panel of reference genomes. The possibility to sequence the whole genome (WGS) using next-generation sequencing technologies (NGS) now allows us to identify rare genetic markers (those present in <1% of the population). Rare variants may usually indicate that mutations at that position modified the gene function, and their low frequency in the population may be due to negative selection. Mutations can involve either gain- or loss-of-function of the gene, which generally have a larger effect on disease risk than SNPs, although they can also exert smaller effects, as SNPs [170].

While the cost of sequencing the genome has fallen from tens of millions to one thousand dollars, it is still prohibitive to do so for larges sample [168]. Selectively sequencing the whole

exome[14] (WES) covers the transcript- and protein-coding regions of the genome, which is also very powerful but with lower costs. The Exome Aggregation Consortium (EXaC) sequenced the exome of 60,706 unrelated individuals from four continents [171]. This project has evolved into the Genome Aggregation Database (gnomAD), which aims to aggregate and harmonize the available exome (n=125,748) and genome (n=15,708) sequencing data [172]. Both WGS and WES enable Rare Variant Association Studies (RVASs) [173].

The combination of high-throughput sequencing approaches and GWASs using genotype imputation enhances the possibilities for new accurate discoveries, as reference panels now have hundreds of thousands of individuals, and genotype imputation will provide a feasible strategy for studying sequencing-discovered variants in millions of genotyped samples [168]. However, the discovery of loss-of-function variants associated with a disease could provide us with very relevant information for identifying potential therapeutic targets [174,175].

- **Genetic biomarkers and coronary heart disease**

Through hypothesis-driven approaches, several genes have been found to be related to monogenic familial CHD, many involved in lipid metabolism. Examples of monogenic CHD-related genes include *MEF2A*, *LPR6*, *CYP27A1* and *ST6GALNAC5* [176]. However, the most fruitful sources of CHD-associated common variants have been GWASs.

[14] The genome sequences that encode functional transcripts and proteins.

The first GWASs identified one locus (chromosome 9p21) as being consistently associated with CHD [166,177–179]. As mentioned above, GWASs have evolved to large international collaborations; in the case of CHD, the biggest one is the CARDIoGRAMplusC4D Consortium (Coronary ARtery DIsease Genome-wide Replication And Meta-analysis plus Coronary Artery Disease Genetics) [180]. This consortium has analysed >200,000 CHD cases and controls of European ancestry and identified 62 loci related to predisposition to CHD that were later validated in a different population [181]. Another recent project is that from the UK Biobank [182], which analysed 34,541 CHD cases and 261,984 controls [183]. Altogether, 163 common variants have been related to CHD at a genome-wide level of significance (Figure 10) [169]. Online public databases with aggregate data on the SNPs-CHD associations are available (www.cardiogramplusc4d.org, biobank.ctsu.ox.ac.uk).

Figure 10. Genes mapping to the CHD risk variants and their pathophysiological pathways in atherosclerosis

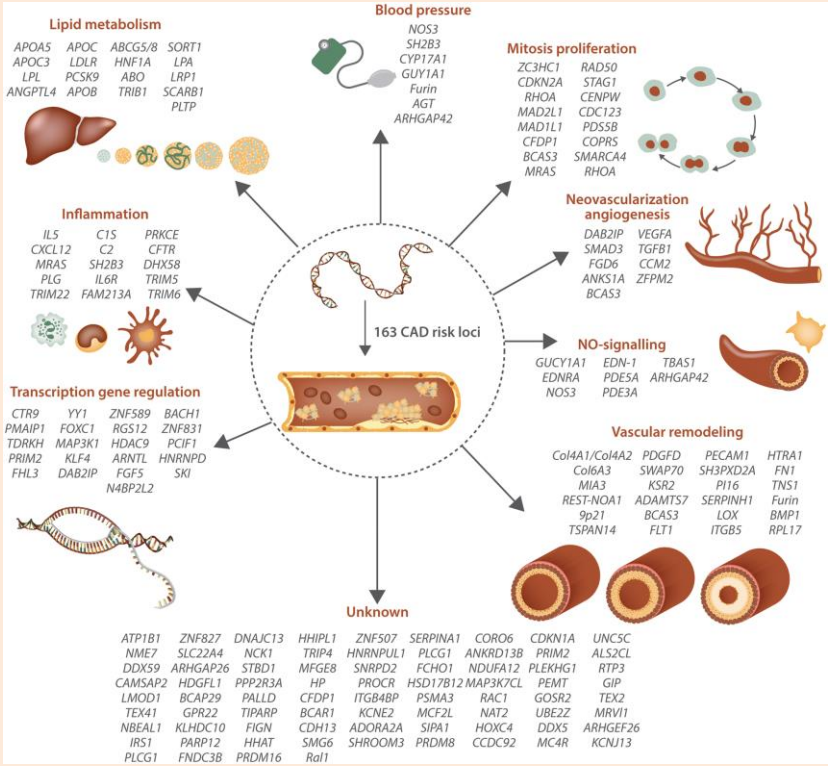


Image from [169].

More than 300 additional SNPs have false discovery rate[15] values below 5%, and are therefore suggestive of CHD risk. Most CHD-related variants have a population frequency of >5%, but have weak effects on CHD risk (~18% increase risk per allele on average) [169,173]. While risk is proportional to the total number of risk SNPs inherited by an individual [184], the combination of these SNPs only explain ~30-40% of CHD heritability (i.e. the proportion of disease variance that is explained by genetic variation) [169]. Some act through CVRFs, e.g. SNPs in *PCSK9*, *LDLR* and *APOE* through cholesterol, and SNPs in *CYP17A1* and *SH2B3* through hypertension [184].

Regarding WGS and WES, still few studies have assessed CHD risk [185]. As expected, the effects of the discovered rare variants on CHD are higher than that of discovered common variants. One study focused on the region containing the CHD-related SNP reported by the first four GWAS (chromosome 9p21) [186]. That region was analysed using NGS and the results compared to those from the pilot of the 1000 Genomes Project, leading to the discovery of rare variants associated with CHD [186]. WES in the gene that encodes adiponectin – which has large effects on families with insulin resistance – also led to the discovery of rare variants associated with CHD [187]. To date, RVASs have found at least nine genes with an aggregation of rare variants that modulate CHD risk. CHD risk was found to be increased in carriers of rare variants within *LDLR*, *LPL,* and *APOA5A*, while carriers of rare variants within *PCSK9, NPC1L1, ASGR1, APOC3, ANGPTL4* and *LPA* had decreased CHD risk [173,188]. NGS efforts will make an important contribution to establishing the bases for precision cardiovascular medicine, as an alternative to the individual prediction strategies currently used in clinical settings [189,190].

Despite the remarkable findings in cardiovascular genetics, the inclusion of individual genetic variants in CVR functions has not sufficiently improved their predictive capacity to be implemented in clinical care. However, combining several weak-effect common

[15] Statistical method to control the proportion of associations that is expected to be false.

variants in Genetic Risk Scores (GRSs), and including these scores in CVR functions resulted in greater improvements in predictive capacity than by including the variants individually [170]. This improvement was greater when using a GRS developed using novel algorithms applied to summary data and imputation from large GWASs in European-ancestry populations. It showed that 8% of the population had greater than threefold increased CHD risk than the general population [191].

- **Deciphering the causal effect of a biomarker on CVD: Mendelian Randomisation studies**

Despite their potential, observational studies neither denote a causal link between the identified exposure (biomarker) and risk of the disease, nor explain how they trigger the disease [170]. Ascertaining the causality of the association between a biomarker and a disease is a condition for defining that biomarker as a potential therapeutic target. While we can correct for some confounding factors (either matching individuals in the design of the study, or adjusting for them in statistical analysis), we cannot discount other potential sources of confounding, and therefore we cannot assume causality in the association. Randomized controlled trials are the optimal approach to assess this, but they are not always possible, and often take years to provide results. Mendelian Randomisation (MR) studies are an alternative that allow researchers to assess this critical issue, using SNPs as proxies for the potentially modifiable biomarker related to population health [192]. They are called "Mendelian Randomisation studies" because they rely on the first two of Mendel's laws:

1. **Law of segregation:** during meiosis, alleles are segregated so that each gamete carries only one allele for each gene.
2. **Law of independent assortment:** the segregation of alleles for one gene occurs independently to that of any other gene.

A MR study has a similar basis to a randomized controlled trial at conception, as the presence of the genetic variant is the only difference between the individuals carrying the biomarker (the intervention group) and those who do not (the control group). Thus, MR analysis is not affected by confounding or reverse causation. However, several assumptions are made [192,193] (Figure 11):

Figure 11. Assumptions of Mendelian Randomisation studies

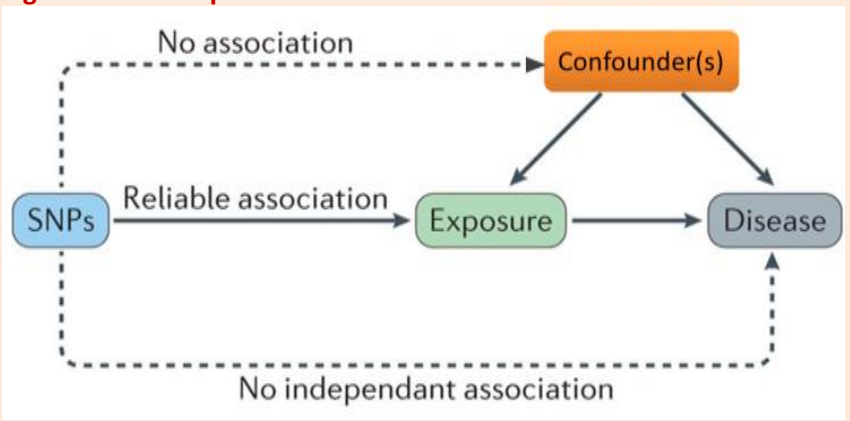


Image modified from [193].

1. **Relevance:** the SNP(s) is/are strongly associated with the biomarker
2. **Independence:** the SNP(s) is/are not dependent on other phenotypes that might act as confounders in the association between the biomarker and the disease
3. **Exclusion restriction:** the SNP(s) affect(s) the outcome only through their effect on the biomarker (lack of horizontal pleiotropy)

While MR methods are currently under development, the basis is well-established. First, one or more SNPs associated with the biomarker are selected to construct the instrumental variable considering the previous assumptions. Usually, at least two SNPs are used to combine (and, thus increase) their effect on the phenotype. Where only one population provides the information about the instrumental variables, the biomarker and the outcome, a single sample MR analysis will be performed. Conversely, if one population provides the association between the genetic instrument and the biomarker, but the association between the same genetic instrument and the outcome is obtained from another population, a two-sample MR analysis will be done. In this case, the instrumental variables must occur in both samples. This approach is highly convenient, as multiple independent MR analyses can be performed using summary results from GWASs as sources, which also provides higher statistical power. However, the two samples must represent the same population with no

overlap among individuals. Another critical issue is horizontal pleiotropy[16], which is not always avoidable. Several methods have been developed for performing a MR study with horizontal pleiotropy, such as the MR Egger regression [192].

If the instrumental variable associated with the biomarker is also associated with the disease, the analysis will support the causal association between the biomarker and the disease. The contrary scenario will suggest that the association between the biomarker and the disease is either in the opposite direction (reverse causation) or mediated by another biomarker. However, it may be also of interest to get information about the magnitude of the causal effect, which can be estimated using the Wald ratio – dividing the effect of the instrumental variable on the outcome by the effect of the instrumental variable on the biomarker [192].

In the case of CHD, several MR studies have confirmed or discarded the causal role of different biomarkers on the disease. Many of the causal biomarkers are CVRFs, such as levels of LDL-C [194] and TG [195], while others are related to inflammation, such as interleukin-6 receptor (IL6R) signalling [196]. Among the biomarkers that showed no causal association with the disease are HDL-C [65], C-reactive protein [197], and adiponectin [198]. However, recent findings suggest that while the circulating levels of a biochemical factor may not be causally linked to the disease, other characteristics of the factor could be. For instance, HDL-C is a controversial factor, as several studies discarded its causal effect in CHD. However, qualitative traits of HDL have been gaining attention in recent years, with some non-peer-reviewed results supporting a causal role of very large HDL particles (available in bioRxiv, DOI: 10.1101/673939). Finally, MR studies are also valuable for predicting the side effects of drugs. For example, one MR study showed that *PCSK9* loss-of-function genetic variants were associated with lower LDL-C levels but also with higher risk of diabetes. This study highlighted the need for a monitored assessment of the participants in trials, or patients treated, with PCSK9 inhibitors to identify this side effect [199].

[16] Horizontal pleiotropy occurs when one variant has independent effects on multiple traits.

1.4. Epigenetics and Epigenomics: a layer of biological information between genetics and lifestyle

As mentioned above, most CHD heritability is not explained by the 164 SNPs found to be associated with CHD. Apart from the rare variants that may be discovered using NGS, some other factors will likely explain this "missing heritability". The term "heritability" refers to the variation in a trait observed between individuals in a given population at a particular time that is attributable to genetic variability in that population. However, heritability does not consider how much of that variability is due to genetic factors [200]. Moreover, the interactions between genes, and between genes and the environment are highly complex and not well-understood yet. Interestingly, alterations in gene expression due to mechanisms other than changes in the DNA sequence are a promising source of missing heritability; they comprise the traits and mechanisms known as "epigenetics".

a) From "Nature versus Nurture" to Epigenetics

Driven initially by scientific curiosity to understand how a zygote evolves, the concept of epigenetics was introduced in 1942 by the embryologist Waddington, who first used "epigenetics" in the context of developmental biology. Some years later he defined it as "the branch of biology that studies the causal interactions between genes and their products, which brings the phenotype into being" [201]. During the following decade, the microbiologist Nanney also referred to "epigenetics", but introduced into its definition the notion of the regulation of gene expression [202]. In the 1960s, gene regulation models were proposed for both bacteria and more complex organisms [203,204], while in the 1970s the component of inheritance was also introduced into the definition of epigenetics. Since then, this concept has evolved and given rise to a research field, with >13,200 English-language articles on human epigenetics published in the last 10 years (PubMed, August 2019). A recently proposed definition states that epigenetics is "the study of molecules and mechanisms that can perpetuate alternative gene activity states in the context of the same DNA sequence" [205].

Epigenetic processes include cell differentiation, phenotypic plasticity[17], genomic imprinting and dosage compensation of the X chromosome. Genomic imprinting refers to a phenomenon that is crucial for normal therian-mammal development in which only one of the two copies of an imprinted gene is functional. It is established in the germ cells by the epigenetic silencing of genes in a sex-specific manner, and is maintained after mitotic division in the somatic cells of the new organism [206]. For instance, the hybrid of a male lion and a female tiger is a liger, whereas that of a male tiger and a female lion is a tigon; the differences in offspring are attributed to imprinting differences among sexes [207]. X-chromosome inactivation is another epigenetic phenomenon that is critical for ensuring gene dosage compensation between sexes in mammals. It consists of random epigenetic silencing of one of the two copies of the X chromosome in the cells of female embryos. Another example, also in felines, is that of a tortoiseshell cat, whose characteristic coat colour is a mosaic of black and orange caused by random X-inactivation in each somatic hair cell [208].

Epigenetic changes and mutations that affect epigenetic components are common in disease. Thus, unravelling whether these changes trigger the disease or play only a "passenger" role will presumably be very valuable for diagnosis, prognosis, and therapy. For instance, several studies show that epigenetic changes are major drivers of oncogenic processes in certain contexts [205].

b) Epigenetic mechanisms

Currently, the most important known carriers of epigenetic information are components of heterochromatin[18], Polycomb proteins, DNA methyl groups, and non-coding RNAs (ncRNAs). The first two are related to histone modifications, which, together with DNA methylation, are the most widely studied epigenetic traits. ncRNAs interact with both histones and DNA methylation, and thereby have a key role in epigenetic regulation as part of a complex

[17] The ability of an individual genome to produce different phenotypes in the presence of different exposures.
[18] Heterochromatin is a tightly packed form of chromatin; see below.

interplay that is not yet completely understood [205]. Given its central role in this thesis, I describe DNA methylation in more detail in the next section.

- **Histone modifications**

DNA is compacted to fit into the cell nucleus as chromatin. Histones are the proteins that package the eukaryotic[19] DNA double-helix into structural units called nucleosomes, which are thus composed of DNA and histones. Nucleosomes, are the basic repeating units of chromatin, and are successively connected by 50 base-pair-long DNA linkers. They consist of an octamer of four core histones (two copies of each of H2A, H2B, H3, and H4) wrapping approximately 147 base pairs of DNA in a little less than two left-handed superhelical turns (Figure 12)

Figure 12. The nucleosome and core histone tail sequences

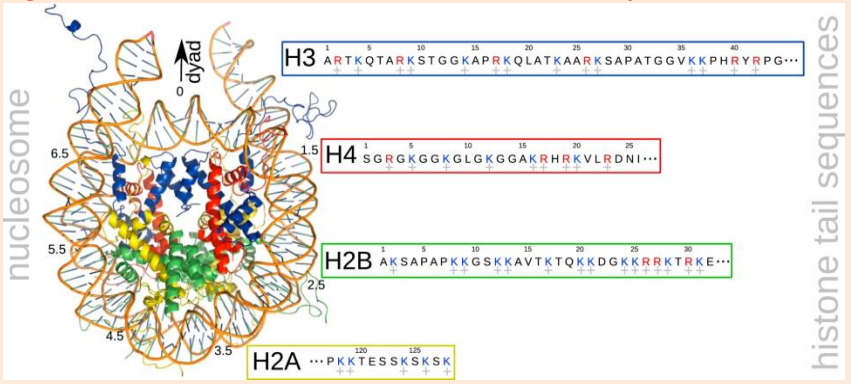


Image from [214].

[209]. Another histone (the linker histone H1) binds at the entry and exit points of the DNA wrapped around nucleosomes and is crucial for more tightly condensing the chromatin and forming the higher order structures [210] then eventually shape a metaphase chromosome (Figure 13). Nucleosomes are not only critical for packaging DNA, but also for ensuring or impairing the DNA sequence's accessibility to proteins involved in DNA replication, recombination, gene expression

[19] Both archaea and eukaryotes express histones, although the structure of the histone-DNA complex is different.

Figure 13. First steps of chromatin folding

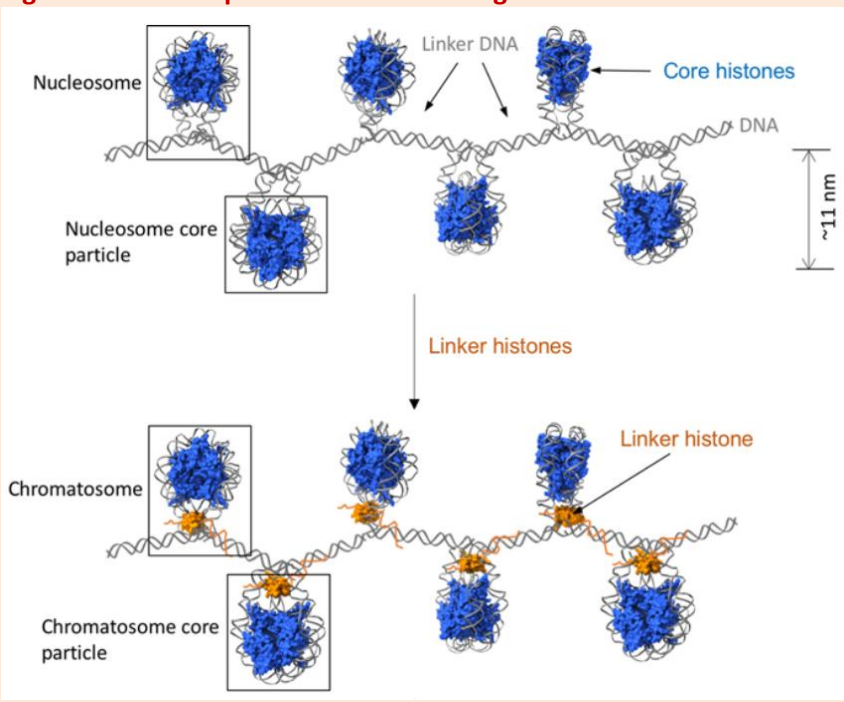


Image modified from [210].

and DNA repair [209]. In this regard, chromatin is locally and reversibly decondensed by several mechanisms, such as ATP-dependent nucleosome remodelling and histone modifications [211].

Histones are rich in positively charged aminoacids, and so are prone to interact with acidic cellular components. Their biochemistry allows them to bind to the negatively charged DNA backbone, although this can result in spurious aggregates. Histone chaperones are proteins that guide nucleosome assembly and control their dynamics, regulating histones from their synthesis, and, in turn, maintaining chromatin homeostasis [212]. However, core histones are characterized by a long tail in their N-terminus – also present in the C-terminus of H2A – that is prone to reversible post-translational modifications [211,213,214].

Histone chemical modifications include acetylation, methylation, phosphorylation, ubiquitilation, sumoylation, ADP-ribosylation, deamination, and proline isomerization, and also their corresponding

reverse mechanisms. In addition to the different classes of reversible modifications that may occur in histone tails, many amino acid residues have been identified for each class; thus, a huge number of combinations can occur simultaneously [209]. Moreover, histone variants can undergo specific modifications at residues that differ from their canonical counterparts [215]. Some histone modifications have been found to be inherited, and most regulate gene expression either by disrupting the contacts between nucleosomes via the histone tails, or by recruiting non-histone proteins. For instance, lysine acetylation – the addition of an acetyl group – is common in actively transcribed regions, but its methylation – the addition of one, two or three methyl groups – can occur in either active or repressed transcription states [209].

- **Non-coding RNAs**

Non-coding RNAs (ncRNAs) are RNA molecules that are not translated into proteins. Transfer RNA (tRNA) and ribosomal RNA (rRNA) are well known to be essential for translating messenger RNA (mRNA, or transcript) into proteins; they are housekeeping ncRNAs. However, there are, presumably thousands of, other RNA molecules that do not encode proteins, and the function of many that have already been discovered remains unknown. One of the functions of the most abundant types is to regulate gene expression at either the transcriptional or post-transcriptional level, as well as epigenetic control. These RNAs mostly act by overlapping genome sequences, and depending on their length they are classified as short (<30 nucleotides) and long ncRNAs (lncRNAs; >200 nucleotides). Recently, circular RNAs (circRNAs) have also been discovered by high-throughput techniques [216–218].

Among the short ncRNAs, microRNAs (miRNAs) and small-interfering RNAs (siRNAs) are involved in the RNA interference (RNAi) pathway, which represses translation by neutralizing target complementary transcripts. The largest type of small ncRNA in animals is the piwi-interacting RNA (piRNA), which forms a complex

with piwi proteins involved in transposon[20] silencing and other functions (Figure 14). Another important type of short ncRNA is small nuclear RNA (snRNA), which is most widely known for establishing an RNA-protein complex that removes the introns of a precursor mRNA to produce a mature transcript during the post-transcriptional process of splicing [216,219].

Figure 14. Processing pathways of some short non-coding RNAs

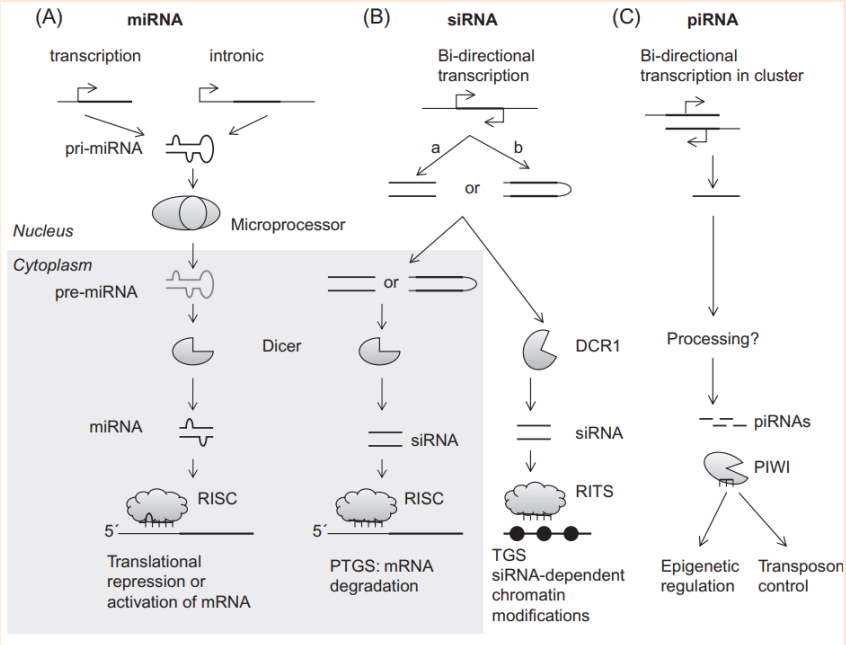


Image from [219].

lncRNAs have not been as well-studied as short ncRNAs, although this is changing thanks to RNA-sequencing, an NGS technique. The most well-known lncRNA is X-inactive specific transcript (Xist), which plays a central role in dosage compensation in human

[20] Transposons are DNA regions that can change its position within the genome, which could result in mutations, more commonly duplications, of the genetic material. They were discovered in maize by Barbara McClintock in the late 1940s, which resulted her being awarded with the Nobel Prize in Physiology and Medicine in 1983 – the only woman ever to be awarded with an unshared Nobel prize in that category.

chromosomes, i.e. the inactivation of one of the two copies of the X chromosome when present [220]. Another promising field of research is that of circRNAs, which are covalently closed single-stranded RNA molecules formed by a specific back-splicing mechanism. They are more stable and resistant to degradation as they lack the 5' terminal cap and the 3' polyadenylated tail. Their functions are still not fully understood, but they have been shown to act as sponges for specific miRNAs that regulate gene expression [221].

c) DNA methylation

DNA methylation is the most widely studied epigenetic mechanism. It consists of the reversible covalent binding of a methyl group to certain sites in the genome without altering the DNA sequence. Although adenine methylation may occur, DNA methylation in eukaryotes typically occurs at the 5th carbon atom in a cytosine ring, which results in 5-methylcytosine (5mC) (Figure 15, [222]) [223,224]. From this point on, when referring to DNA methylation, I am referring to 5mC.

Figure 15. Scheme illustrating methylation at cytosine residues

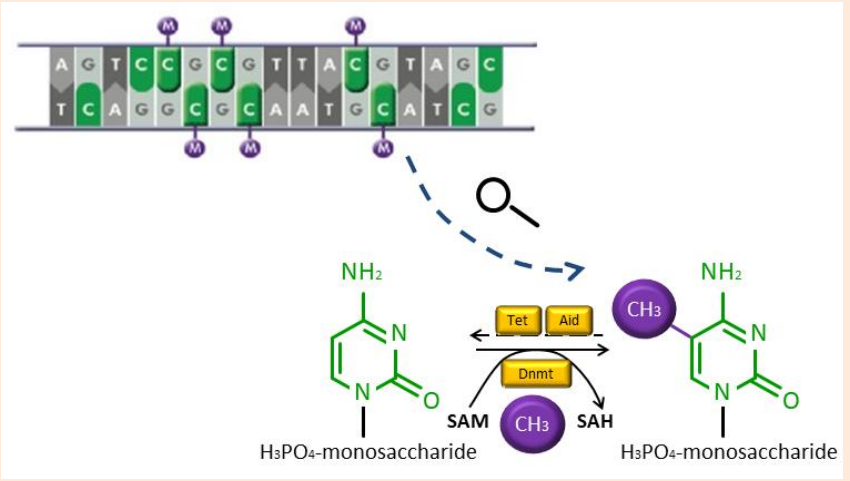


Image adapted from [222].

- **DNA methylation sites**

This epigenetic mechanism mostly occurs at cytosines that are followed by guanines, a dinucleotide known as a CpG site. While more than 70% of CpG sites are methylated in vertebrate somatic tissues in general, this dinucleotide is quite rare in mammalian genomes, which have 5-fold fewer CpG sites than expected from their nucleotide composition. This is due to the mutagenic state of 5mC, which can spontaneously deaminate into thymine. In fact, this mechanism has mostly been lost in other animals such as *Drosophila melanogaster, Caenorhabditis elegans*, and fission and bakers' yeasts [223,224].

More recently, studies in humans have reported methylation at cytosines that are followed by nucleotides other than guanine – non-CG methylation, mCH, where H is A, C, or T, and have found significant numbers of these methylated dinucleotides in pluripotent and brain cells. Also, more mCH were observed in the promoter of *PGC-1a* in diabetic than in healthy individuals. However, the biological function and molecular mechanisms involved in generating mCH are still unclear. [225].

- **DNA methylation writers and erasers**

DNA methyltransferases (DNMTs) are the enzymes that transfer the methyl group from the cofactor S-adenosyl-L-methionine (SAM) to the cytosine. *De novo* DNMTs establish DNA methylation marks at previously unmethylated sites, while maintenance DNMTs preserve established DNA methylation marks during DNA replication. In mammals, this last process is catalysed by one member of the Dnmt1 family accompanied by the UHRF1 ligase, while *de novo* methylation is mediated by both the Dnmt3a and Dnmt3b enzymes, which use DNMT3L cofactor in germ cells. While the specific mechanisms that target Dnmt3 to certain DNA regions are still not fully understood, ncRNAs and histone modifications have been linked in this epigenetic interplay [223,224].

DNA methylation is a dynamic process that is achieved by two mechanisms, DNA replication-dependent passive demethylation, and enzyme-mediated active demethylation. The latter is mediated by TET methylcytosine dioxygenases, which progressively oxidize 5mC to 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC) [223].

- **Functions of DNA methylation**

The main functions of DNA methylation are related to the fact that at least 70% of vertebrate CpGs are methylated. Most of the genome consists of repetitive elements that are enriched in CpG sites. In fact, transposons and satellite repeats that occur close to centromeres and telomeres are highly methylated. Consequently, DNA methylation in these regions confers stability to the genome and preserves its integrity by avoiding gene disruption [223,224]. Silencing of transposons has been related to ncRNA [223].

Moreover, the promoter regions of genes are enriched in CpG sites – regions called CpG islands or CGIs in mammals – and this feature is critical for regulating gene expression and silencing. Promoters of active housekeeping genes are mostly non-methylated (hypomethylated), allowing transcription factors binding [223,224]. Conversely, the coding regions of active genes are usually methylated (hypermethylation), although it is unclear why. It has been proposed that this feature facilitates transcription elongation and co-transcriptional splicing, and that it may repress intragenic cryptic promoters. Both hypo- and hypermethylation at promoters and gene bodies, respectively, are related to the level of accessibility of the DNA to the DNMTs as a consequence of histone modifications produced following gene transcription. Similarly, gene silencing results from the interplay between DNA methylation-related binding proteins and chromatin remodellers and modifiers [223].

Note that there are some exceptions of the general hypomethylated state of gene promoters. Stable and lifelong promoter silencing by DNA methylation is key for three critical biological processes: X-chromosome inactivation, genomic imprinting and germline-cell specification [223].

- **DNA methylation reprogramming**

During mammalian development, the epigenome needs to be reprogrammed to a totipotent state for the next generation, and epigenetic marks are thus erased and remodelled. The epigenome undergoes two extensive waves of global demethylation and remethylation: one after fertilization and one during gametogenesis.

Some exceptional sites can evade demethylation, sites that are related to imprinting and transposon repression, and this fact explains epigenetic inheritance [223,226].

In the fertilized zygote, the paternal and maternal genomes undergo active – mediated by TET3 – and passive demethylation, respectively, until the blastocyst stage. After blastocyst implantation, *de novo* DNA methylation establishes the initial embryonic methylation pattern [223,226].

In the epiblast, a subset of stem cells is specified for the germline, and the primordial germ cells undergo a two-step DNA demethylation process, first passive, then mediated by TET1 and TET2. Then, sex-specific germ-cell methylation patterns are established. In the case of male gametes, DNMT3A and DNMT3L methylate the genome before birth. Conversely, the oocyte genome is methylated after meiosis and prior to ovulation through the activity of DNMT3A [223,226].

d) Assessing DNA methylation

Many different methodologies are available for analysing DNA methylation, the specific choice depending on the specific biological question. Broadly, DNA methylation can be assessed at a global level or at DNA methylation sites, either at specific genes or genome-wide. Regarding the latter, it is more accurate to use the term "epigenome-wide", as epigenetic marks constitute another layer of biological information that forms the so-called "epigenome" and analogous with the "genome". Similarly, the DNA methylation marks of a single genome constitute a "methylome". Therefore, this epigenome-wide approach to assessing DNA methylation allows researchers to decipher methylomes. The gold-standard techniques are based on bisulfite treatment of the DNA, which is the basis of high-throughput technologies for methylome profiling (see Box 5).

Box 5. Basic concepts in genomic technologies

- **Probe**, also known as oligonucleotide, is a short single-strand DNA (or RNA) molecule that is complementary to a specific sequence
- **PCR,** a basic method to amplify copies of DNA fragments in successive temperature-dependant cycles consisting of denaturation, annealing, and elongation steps
- **Genomic restriction digest**, a basic method for obtaining DNA fragments using restriction enzymes, which cleave DNA in a sequence-specific manner
- **DNA sequencing**, a process of determining the nucleotide sequence using different methods, such as Sanger sequencing or pyrosequencing
- **Array**, a collection of specific probes organized in a grid of microscopic spots attached to a solid surface

• Approaches to assessing DNA methylation

As mentioned, DNA methylation can be assessed at a global level, or at DNA methylation sites. The first approach does not give information about specific alterations in DNA methylation patterns but rather about overall changes, while the other allows one to identify differential methylation patterns that may be linked to a functional outcome.

▪ Global DNA methylation

Several techniques have been developed for profiling whole genome methylation, but only High-Performance Liquid Chromatography (HPLC) and bisulfite-based Polymerase Chain Reaction (PCR) methods have been found to provide correlated results. The first is based on hydrolysing DNA to separate and identify the deoxyribonucleosides according to their ultraviolet absorbancies. The ratio of 5mC/dC can be calculated for each sample. Bisulfite-based PCR methods used to assess overall DNA methylation are based on conservative repetitive elements, mainly the long interspersed nuclear element 1 (LINE-1). They involve bisulfite treatment (see below) prior to PCR amplification of those elements, which are then quantified by pyrosequencing [227].

- **DNA methylation sites**

Differentially methylated sites can be assessed at specific genes or epigenome-wide. The most common methods for the first approach can be grouped into sensitive or quantitative approaches. One sensitive method is the Methylation-Specific PCR (MSP), which is based on two sets of primers designed to amplify either methylated or unmethylated alleles. This is a rapid, highly sensitive technique that is suitable for samples with limited quantity and quality. Quantitative methods are based on either methylation-sensitive restriction analysis, or bisulfite conversion of DNA. The first methods are based on the selective digestion of DNA by endonucleases (HspI and MspI) [227], while in the latter, both methylated and unmethylated alleles are amplified from bisulfite-modified DNA, and the subsequent techniques vary in the different methods used [228].

Thanks to technological advances, most current DNA methylation research now uses hypothesis-free approaches. Several methods are available based on different techniques (genome restriction, immunoprecipitation and bisulfite conversion), with the most popular ones using microarray and NGS technologies based on bisulfite treatment of the genome. Whole Genome Bisulfite Sequencing (WGBS) is probably the most powerful and comprehensive method for assessing methylomes, and is similar to WGS but with the additional step of bisulfite conversion. However, its cost and complex data analysis are limiting factors; moreover, many studies do not require the whole genome, as not all is methylated. An alternative to WGBS is Reduced Representation Bisulfite Sequencing (RRBS), which uses restriction enzymes to obtain CpG-enriched fragments of the genome for further sequencing [227,229]. Arrays based on Illumina probe extension assays – the Infinium HumanMethylation450 and Methylation EPIC BeadChips – are the most extensively used methods for methylome profiling, as they enable one to assess multiple samples at a lower cost, and the data analysis is less complex [229].

- **Bisulfite conversion-based arrays**

Bisulfite conversion methods consist of bisulfite treatment of DNA before sequencing to determine DNA methylation patterns at single-nucleotide resolution. Bisulfite treatment deaminates unmethylated cytosines to uracils, which are not present in DNA and so are converted into thymines in the subsequent analysis. 5mC residues are resistant to bisulfite treatment, so they remain as cytosines, and can thus be detected by comparing bisulfite-treated and untreated samples [227].

Microarrays based on bisulfite conversion use locus-specific PCR to interrogate hundreds of thousands of chemically differentiated cytosines simultaneously [227]. The most common arrays are commercialized by Illumina: the Infinium HumanMethylation450 and Methylation EPIC BeadChips (hereafter, the 450k array and EPIC array, respectively). They use the Infinium Methylation Assay chemistry, which is subdivided into the Infinium I and II Assays depending on the number of specific probes used per methylation site. The Infinium I Assay uses one probe that is specific to the methylated site and another for the unmethylated locus, and was the only assay used in the first Illumina Methylation array, the Infinium HumanMethylation27 BeadChip. In contrast, the Infinium II Assay requires only one probe per locus. The combination of both chemistries enhances the epigenome coverage of the array (Figure 16) [230].

The 450k array covers more than 485,000 methylation sites (482,421 CpG sites, 3,091 non-CpG sites and 65 random SNPs) [231]. While many studies have used this array, it is now no longer commercialized. The more recent EPIC array interrogates 853,307 CpG sites, including 91.1% of the 450K probes [232]. The content of the arrays was selected according to the recommendations of a worldwide consortium. Probes were randomly assigned to wells in the array – whose surface is covered with silica beads – so that each is represented by 15-30 beads, which provides multiple measurements for each locus [231].

Figure 16. Scheme illustrating the Infinium Methylation Assay

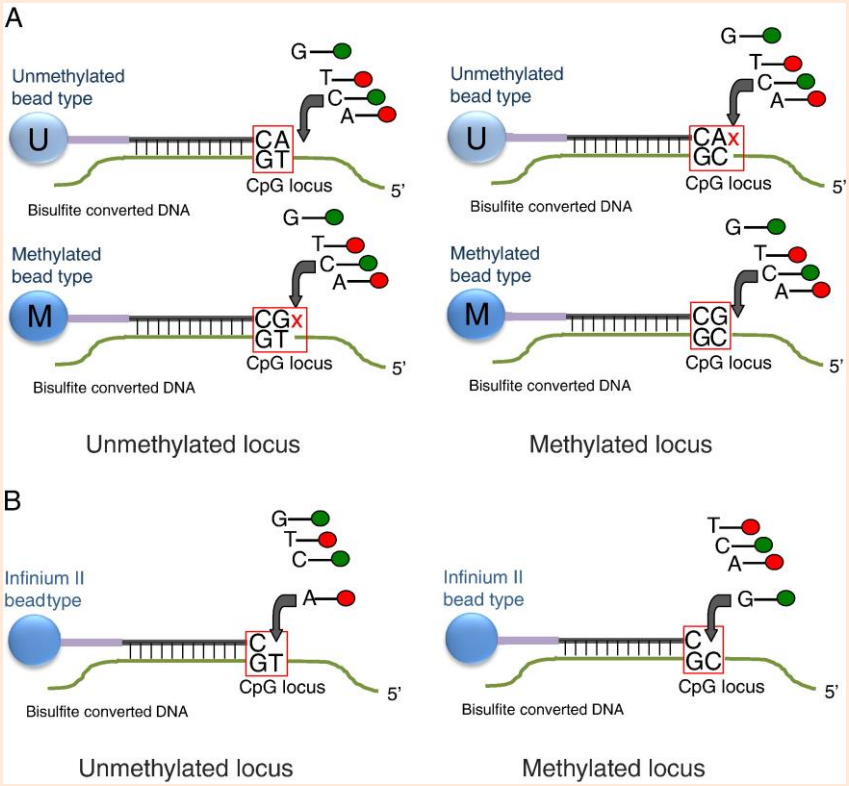


Image from [231].

The first steps of the Infinium assay (amplification of bisulfite-treated DNA samples up to 1,000-fold; endpoint DNA fragmentation into 300-600 base-pair fragments; and precipitation and resuspension – purification – of the DNA fragments) are performed in microplates, with one sample per well. The samples are then transferred to the array where they bind to locus-specific probes attached to spherical beads embedded in the array surface (hybridization step). Then, single-labelled base extension of the probes and immunohistochemistry staining (XStain step) take place simultaneously. Fluorescent signals are imaged on an Illumina scanner using red and green lasers, which generates fluorescence intensity data consisting of two files per methylation locus per sample, one each for the red and green intensity data [230,231].

Typically, fluorescence data are then processed using bioinformatics resources and tools. Quality control and normalization of the data are required prior to obtaining a methylation value for each locus and each sample [233–235]. The level of methylation for each locus can be determined as the ratio between the fluorescence signals of the methylated and unmethylated sites. Methylation values for subsequent analyses can be expressed as M- or β-values, each of which can be converted to the other [236] (see Box 6).

Box 6. How to express methylation values

- **M-value,** positive values indicate the presence of more methylated than unmethylated cytosines, while negative values denote the opposite. The M-value is more statistically robust than the β-value.

$$M\ value = \log_2\left(\frac{M_i + \alpha}{U_i + \alpha}\right)$$

- **β-value,** range between 0 (completely unmethylated) and 1 (completely methylated). The β-value provides a more straightforward interpretation than the M-value.

$$\beta\ value = \frac{M_i}{M_i + U_i + \alpha}$$

In the equations:

M_i = intensity of methylated probes,

U_i = intensity of unmethylated probes, and

α = constant offset (α=1 and α=100 in the equations for the M- and β-value, respectively).

e) Identification of DNA methylation biomarkers

As for genetic biomarkers, DNA methylation patterns can be used as biomarkers of health outcomes, including cardiovascular diseases, leading to the emergence of epigenetic epidemiology in the past decade. As in genetic epidemiology, association studies are currently the most common approaches for identifying methylome patterns related to diseases, exposures or lifestyle/environmental factors. Differentially methylation sites can be found using either a hypothesis-driven approach, i.e. selecting specific loci, or using a hypothesis-free

approach, i.e. assessing the methylome. The latter is known as an Epigenome-Wide Association Study (EWAS), although this term should indeed encompass other epigenetic traits. Currently, most epigenetics research groups use this approach, and have established similar collaborations to those that have characterised GWASs research.

• Epigenome-wide association studies

EWASs assess the association between a phenotype of interest and the methylation status of cytosines distributed throughout the genome, or more precisely, those that are covered by the available commercial arrays. Two databases with information on published EWASs are available online: the EWAS catalog (http://www.ewascatalog.org/) and the EWAS Atlas (http://bigd.big.ac.cn/ewas/index). When the phenotype of interest is a disease, these resources can improve our understanding of its aetiology and provide new targets for therapeutic and prediction purposes. Because of its dynamic chemical nature, DNA methylation can be altered throughout life by environmental and lifestyle factors, so EWAS results are very informative about the interplay between environment/lifestyle, genes, and health status. However they are only valid for the particular time point at which samples were collected. Consequently, these studies cannot always infer the causality of the association, i.e. changes in DNA methylation patterns may be either the cause or a consequence of the disease, so it will generally be necessary to conduct MR studies and to integrate EWAS and GWAS (and ideally transcriptional or proteomic data). Other important considerations include the tissue selected for collecting the DNA sample, and statistical issues [237–242].

Blood is the most common source sample for DNA methylation profiling. Other tissues can be also processed, although generally only one tissue is analysed per study. However, again due to DNA methylation specificity and dynamics, not all tissues of an organism show the same DNA methylation patterns, nor do all cells. Several studies have suggested that blood provides a good proxy for methylation in other tissues [237–239]. To overcome cell-type confounding, Houseman *et al* developed an algorithm based on reference methylation data to estimate the proportion of different cell types in the analysed sample – valid for blood, umbilical cord and some brain cells. For instance, the proportions of T lymphocytes, B

lymphocytes, monocytes, Natural Killer lymphocytes and neutrophils are inferred for blood samples [243]. Subsequently, other algorithms were also developed, some of which do not require any reference methylation profiles [241,244].

EWASs can also be confounded by population characteristics or by systematic differences between the compared individuals due to sample processing (known as batch effect). It is highly recommended that the study design and analysis include potential confounders such as age, sex, ethnic background, and smoking behaviour [237]. Moreover, high-throughput data allow us to correct for confounding using estimated covariates such as principal components or surrogate variables [237,245].

In addition, to obtain results with high statistical power, large samples are essential. However, EWAS usually has higher statistical power than GWAS for the same sample size, since EWAS uses a continuous variable instead of a categorical one [237,239]. Also, since each analysis performs multiple comparisons, to reduce the number of false positives the results must be corrected for multiple testing using the Bonferroni or the false discovery rate (FDR) corrections [240].

The nature of EWAS require the results to be validated [237–240]. As in GWAS, EWAS design typically consists of two stages: a discovery and a validation stage performed in independent populations. The most commonly used strategy is to initially tolerate some false positive results in the discovery stage, and to include and analyse them in the validation stage [237,239,240]. Additionally, meta-analysing multiple studies boosts the total sample size [237,240].

In addition to the differentially methylated cytosines reported by EWAS, more recently there has been increasing interest in the identification of Differentially Methylated Regions (DMRs). Methylation levels are strongly correlated throughout the genome, and the functionally relevant findings that have been reported have generally been associated with genomic regions. These analyses may increase the statistical power of EWAS, and different methods have been developed for analysing DMRs, either by aggregating single sites or direct methods [239,241,242,246].

- **DNA methylation as a source of biomarkers of cardiovascular diseases and related factors**

A famous unplanned "experiment" was the Dutch famine of 1944-45, in which individuals exposed to malnutrition *in utero* and their direct descendants in adulthood had higher rates of obesity and hypertension, increased CVR, and impaired glucose metabolism. They were subsequently found to have different levels of methylation in genes related to those outcomes [247]. Conversely, individuals who were exposed to malnutrition at a more advanced gestational stage only had different levels of methylation in one of those identified genes [248]. This is an example of how environmental or behavioural factors can modulate cardiovascular-related outcomes via methylation changes [242].

Most classical CVRFs have been associated with differential methylation patterns in EWAS analyses (http://www.ewascatalog.org/; http://bigd.big.ac.cn/ewas/index), and there is ongoing research into their biological and clinical relevance. For instance, epigenetic differences in monozygotic twin pairs accumulate with age, with larger differences in older twin pairs [249]. In fact, age is highly related to epigenetics, to the point that the concept of "epigenetic age" has been proposed, in which age estimators are based on methylation loci ("epigenetic clocks") [250]. Another widely reported factor that alters DNA methylation is smoking exposure. The *AHRR* and *F2RL3* genes have been shown to be differentially methylated at high levels in different smoking contexts [251,252], and methylation at the latter was linked to mortality in patients with stable CHD, suggesting a potential mechanistic role for DNA methylation.[253].

As observed in other complex diseases [254], current research supports the notion that changes in DNA methylation play an important role in susceptibility to CVD and in the development of atherosclerosis [252]. There is growing evidence to suggest that epigenetic modifications actively reshape pathological processes in CVD, such as the dedifferentiation of smooth muscle cells or the accumulation of senescent cells [255]. The pathogenesis of atherosclerosis also involves dynamic epigenetic changes in a cell type- and stage-specific manner, which has motivated efforts to decipher them for both therapeutic and biomarker purposes [256]. Also,

integrating DNA methylation (ideally epigenetic data) and genetic information may provide information for improved CVD prediction, as an alternative to classical CVRFs [257]. However, it is still not clear whether DNA methylation is a major driver of the pathogenesis of CVD, and it cannot be ruled out that methylation changes are a consequence of CVD, rather than a cause [242,249,258].

1.5. Justification of this thesis

Despite extensive knowledge of the health impacts of some CVRFs, CVD, and specifically CHD, are still the leading cause of mortality and morbidity worldwide. In addition, the genetic component of CHD that has been identified so far does not completely account for the interindividual variation that is not explained by its environmental or lifestyle components. Thus, for a better understanding of CVD biology, further studies are required to address the source of biological inheritance, DNA, and how this genetic information is influenced by the environment.

The layer of biological information that modulates the expression and stability of genomes is the Epigenome, which consists of heritable traits regulating gene expression that are independent of DNA sequence variation. Additionally, some of these epigenetic traits can be modified by physiological, behavioural or environmental factors. In this regard, epigenetic patterns may partly explain the missing heritability of CVD, they may be a biomarker of the so-called exposome (environment + lifestyle), and may have an intermediate and mediating role between cardiovascular risk factors and CVD risk. More specifically, I will focus on the most widely studied epigenetic mechanism, DNA methylation.

On the one hand, a hypothesis-free approach to studying DNA methylation throughout the genome – the methylome – has the potential to discover new mechanisms or traits underlying CVD. On the other hand, a hypothesis-driven approach based on previous knowledge allows us to elucidate the precise mechanisms that link the candidate genes to the disease. Together, these approaches could provide us with new mechanisms and predictive biomarkers for CVD and deeper knowledge of the role of DNA methylation in CVD. Finally, assessing the causal relationship between a biomarker and a disease will provide us with information about the potential role of this biomarker as a therapeutic target for preventing or treating that disease.

In this thesis I will address some of these issues to contribute to disentangling this complex scenario:

a.- Physical activity has been consistently reported in cohort studies to be inversely associated with CVD risk. However, the molecular mechanisms and pathways that underlie this association are still not fully understood. Thus, the first objective of this thesis is to assess the association between physical activity and DNA methylation using an unbiased, hypothesis-free approach. This study could provide us with new insights on molecular mechanisms that may explain the health benefits of physical activity.

b.- The accumulation of CVRFs exponentially increases a person's risk of CVD, and this accumulation could activate molecular mechanisms that are inactive when only isolated risk factors are present. In this context, the second objective of this thesis is to assess the association between the total burden of CVRFs (independently of age) and DNA methylation using a hypothesis-free approach. This analysis could provide us with new molecular mechanisms related to the accumulation of CVRFs that may have an impact on CVD risk.

c.- There is limited evidence on the association between epigenetic signatures and cardiovascular health outcomes. In this thesis, I will address the relationship between DNA methylation and CVD risk. First, I will perform a systematic review and meta-analysis to summarize current knowledge about this relationship. Second, I will use a hypothesis-free (EWAS) and candidate gene approach to identify DNA methylation biomarkers associated with cardiovascular outcomes (AMI, CHD and general CVD). I will also assess the clinical utility of the identified CVD-related CpGs as predictive biomarkers for CVD by including them in a CVR function. Finally, I will tackle the causality of those epigenetic-CVD associations using a Mendelian Randomisation approach.

2. HYPOTHESES AND OBJECTIVES

Illustration designed by "Vectorarte / Freepik"

"The most difficult thing is the decision to act, the rest is merely tenacity."

Amelia Earhart (1897–1939).

2.1. Hypotheses

DNA methylation is a dynamic chemical modification of the nitrogenous base of a nucleotide. Thus, because of its reversible nature, it could partly mediate the effects of lifestyle and environmental factors on the risk of CVD and CHD. In addition, DNA methylation patterns are heritable, so they could partly explain the missing heritability and lead to the discovery of new pathogenic mechanisms underlying these complex diseases.

a) Hypothesis 1

- **General hypothesis**

Lifestyle and environmental factors influence DNA methylation.

- **Specific hypothesis**

Physical activity has multiple health benefits, including for cardiovascular health. These benefits may be related to differential levels of DNA methylation.

b) Hypothesis 2

- **General hypothesis**

DNA methylation patterns are associated with individual CVRFs. The accumulation of CVRFs, independently of age, could also lead to specific DNA methylation signatures.

- **Specific hypotheses**

b.1) CVR related to the accumulation of individual and classical CVRFs influences DNA methylation, independently of age.

b.2) This DNA methylation signature is associated with subclinical atherosclerosis and with the incidence of cardiovascular events.

c) Hypothesis 3

• General hypothesis

DNA methylation could partly determine the occurrence of CVD and CHD events.

• Specific hypotheses

c.1) DNA methylation patterns may be associated with both prevalent and incident cases of CVD and CHD.

c.2) DNA methylation patterns associated with cardiovascular events may be related to classical CVRFs.

c.3) DNA methylation patterns associated with cardiovascular events may be biomarkers of cardiovascular and coronary risk.

- Risk scores based on DNA methylation patterns associated with cardiovascular events may predict future events, independently of classical CVRFs.

- These risk scores may improve the predictive capacity of classical CVR functions.

c.4) Some of the DNA methylation marks identified as being associated with cardiovascular events are causally related to these complex outcomes.

2.2. Objectives

The main objective of this thesis is to use population-based studies to identify individual CpGs throughout the genome that show differential levels of methylation in association with cardiovascular and coronary risk.

a) Objective 1

- **General objective**

To identify differential DNA methylation patterns associated with a lifestyle factor: physical activity.

- **Specific objective**

To identify DNA methylation loci that are related to leisure-time physical activity by performing a two-stage epigenome-wide association study (***Manuscript 1***).

b) Objective 2

- **General objective**

To identify differential DNA methylation signatures that are associated with the accumulation of classical CVRFs.

- **Specific objectives**

b.1) To identify differential DNA methylation loci related to CVR by conducting a two-stage epigenome-wide association study of age-independent CVR (***Manuscript 2***).

b.2) To assess whether these DNA methylation signatures are associated with subclinical atherosclerosis and with the incidence of cardiovascular events (***Manuscript 2***).

c) Objective 3

• General objective

To identify differential DNA methylation patterns associated with CVD and CHD and to evaluate their potential as predictive cardiovascular biomarkers.

• Specific objectives

c.1) To identify differential DNA methylation loci related to CVR by performing a systematic review of evidence on this topic (***Manuscript 3***).

c.2) To identify differential DNA methylation loci related to prevalent cases of CHD and incident cases of CVD and CHD using both a candidate-gene and an epigenome-wide strategy in population-based studies (***Manuscript 4***).

c.3) To assess the associations between these loci and classical CVRFs in population-based studies (***Manuscript 4***).

c.4) To determine whether these loci are suitable potential predictive biomarkers of CVD and CHD (***Manuscript 4***).

- To construct risk scores based on those loci and assess their association with future CVD and CHD events in a prospective cohort study.

- To include those risk scores in a CVR function to evaluate in a prospective cohort study whether this improves the function's capacity to estimate cardiovascular or coronary risk compared to the classical CVR function.

c.5) To assess the causality of the associations between the identified DNA methylation loci and CHD using Mendelian Randomisation studies (***Manuscript 4***). .

3. MANUSCRIPTS

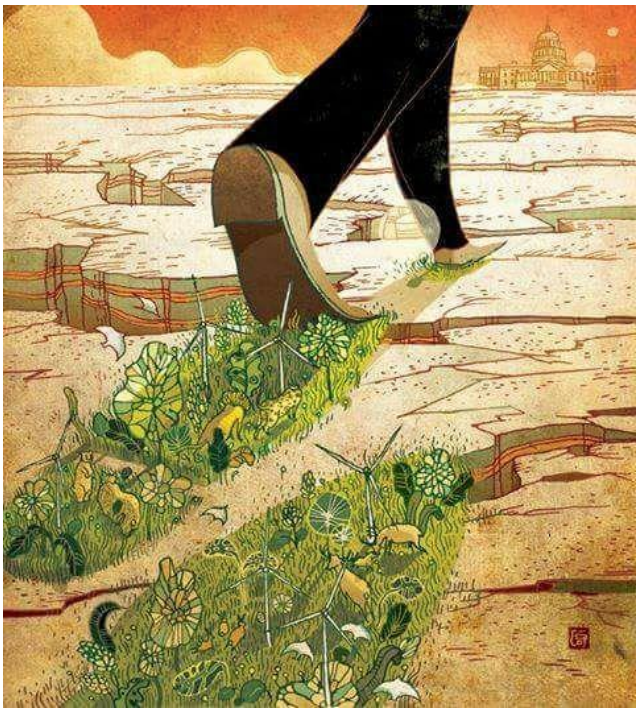
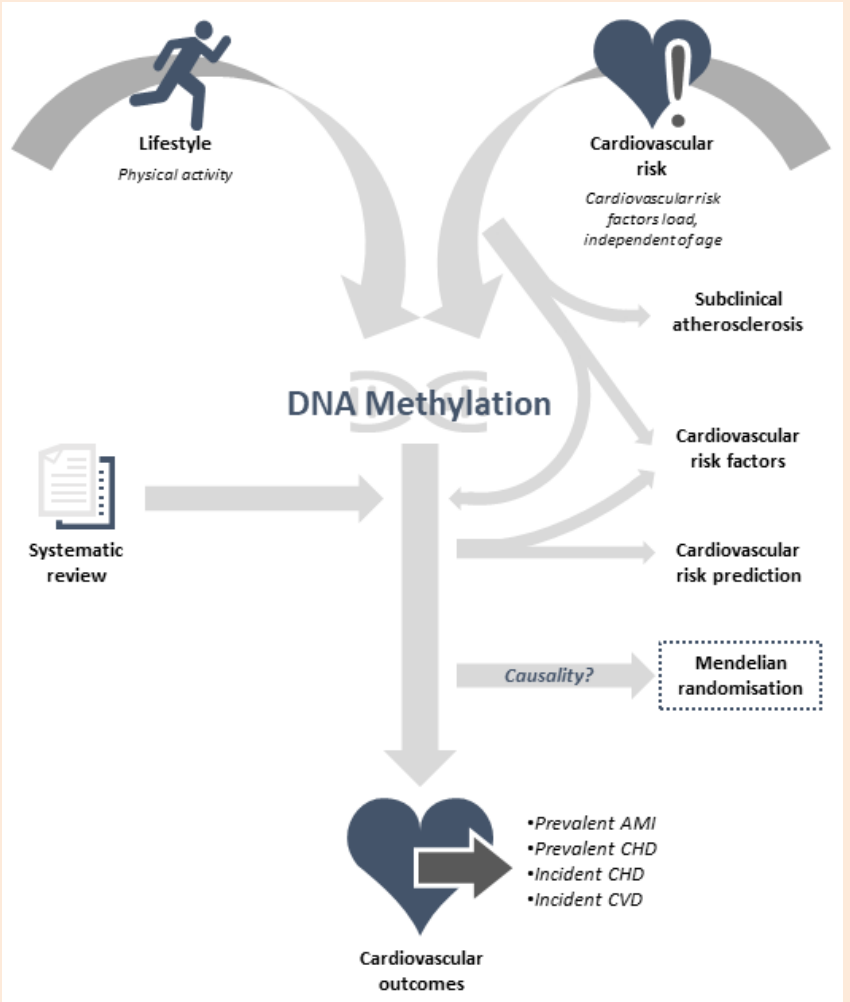


Illustration by Victo Ngai

"I was taught that the way of progress was neither swift nor easy."

Marie Curie (1867-1934).

Figure 17. Graphical abstract of the analysis performed in this thesis



** AMI, Acute Myocardial Infarction; CHD, Coronary Heart Disease; CVD, CardioVascular Disease.*

Table 1. Summary of the methods used in the main analyses of this thesis.

Objective – Manuscript	Study design	Cohorts	Exposure	Outcome	Statistical approach
Objective 1 – Manuscript 1	• Cross-sectional Two-stage EWAS	• REGICOR • FOS	PA questionnaires; • LPA, MPA, VPA • LPA, MVPA • TPA	Blood methylomes; • 450k • EPIC	• Generalized additive models • Metaanalysis
Objective 2 – Manuscript 2	• Cross-sectional Two-stage EWAS	• REGICOR • FOS	CVRFs load excluding age; • Δvascular age-chronological age • Residuals from age/CVR relation	Blood methylomes; • 450k	• Robust linear regression • Cox regression • Metaanalysis
Objective 3 – Manuscript 3	• Systematic review	–	DNA methylation - CVR		• Probability of finding a CpG/ gene in ≥2 studies by chance
Objective 3 – Manuscript 4	• Age- and sex-paired case-control EWAS • Prospective EWAS • Nested case-control EWAS • Metaanalysis of 2 EWASs	• REGICOR • FOS • WHI	Blood methylomes; • 450k • EPIC	• Prevalent AMI • Prevalent CHD • Incident CHD • Incident CVD	• Logistic regression • Cox regression • Metaanalysis

**EWAS, Epigenome-Wide Association Study; FOS, Framingham Offspring Study; WHI, Women's Health Initiative; PA, Physical Activity; LPA, Light PA; MPA, Moderate PA; VPA, Vigorous PA; MVPA, Moderate-Vigorous PA; TPA, Total PA; CVRF, CardioVascular Risk Factor; CVR, CardioVascular Risk; 450k, Infinium HumanMethylation450 BeadChip; EPIC, Infinium Methylation EPIC BeadChip; AMI, Acute Myocardial Infarction; CHD, Coronary Heart Disease; CVD, CardioVascular Disease.*

3.1. Manuscript 1

Fernández-Sanlés A, Sayols-Baixeras S, Castro de Moura M, Esteller M, Subirana I, Torres-Cuevas S, Pérez-Fernández S, Aslibekyan S, Marrugat J, Elosua R. Physical activity and genome-wide DNA methylation: the REGICOR study. Medicine & Science in Sports & Exercise. In press.

Fernández-Sanlés A, Sayols-Baixeras S, Castro De Moura M, Esteller M, Subirana I, Torres-Cuevas S, et al. Physical Activity and Genome-wide DNA Methylation: The REgistre GIroní del COR Study. Med Sci Sports Exerc. 2020 Mar 1;52(3):589–97. DOI: 10.1249/MSS.0000000000002174

Supplemental information is provided in the attached CD and will be available soon online at the journal's website (see next pages).

3.2. Manuscript 2

Fernández-Sanlés A, Sayols-Baixeras S, Curcio S, Subirana I, Marrugat J, Elosua R. DNA Methylation and Age-Independent Cardiovascular Risk, an Epigenome-Wide Approach: The REGICOR Study (REgistre GIroní del COR). Arterioscler Thromb Vasc Biol. 2018;38(3):645-52.

Fernández-Sanlés A, Sayols-Baixeras S, Curcio S, Subirana I, Marrugat J, Elosua R. DNA methylation and age-Independent cardiovascular risk, an epigenome-Wide approach the REGICOR study (REgistre GIroní del COR). Arterioscler Thromb Vasc Biol. 2018 Jan 1;38(3):645–52. DOI: 10.1161/ATVBAHA.117.310340

Supplemental information is provided in the attached CD and is available online at the journal's website.

3.3. Manuscript 3

Fernández-Sanlés A, Sayols-Baixeras S, Subirana I, Degano IR, Elosua R. Association between DNA methylation and coronary heart disease or other atherosclerotic events: A systematic review. Atherosclerosis. 2017;263:325-33.

Fernández-Sanlés A, Sayols-Baixeras S, Subirana I, Degano IR, Elosua R. Association between DNA methylation and coronary heart disease or other atherosclerotic events: A systematic review. Atherosclerosis. 2017; 263:325-333. DOI: 10.1016/ j.atherosclerosis.2017.05.02.

Supplemental information is provided in the attached CD and is available online at the journal's website.

3.4. Manuscript 4

Fernández-Sanlés A, Sayols-Baixeras S, Subirana I, Sentí M, Subirana I, Pérez-Fernández S, Castro de Moura M, Esteller M, Marrugat J, Elosua R. DNA methylation biomarkers of myocardial infarction and cardiovascular disease**.** bioRxiv. 2019.

Fernández-Sanlés A, Sayols-Baixeras S, Subirana I, Sentí M, Pérez-Fernández S, Castro de Moura M, et al. DNA methylation biomarkers of myocardial infarction and cardiovascular disease. bioRxiv. bioRxiv; 2019. p. 707315. DOI: 10.1101/707315

Supplemental information is provided in the attached CD and is available online in bioRxiv.

4. DISCUSSION

Illustration by Gérard Dubois

"Only when our clever brain and our human heart work together in harmony can we achieve our true potential."

Jane Goodall (1934).

4.1. General overview

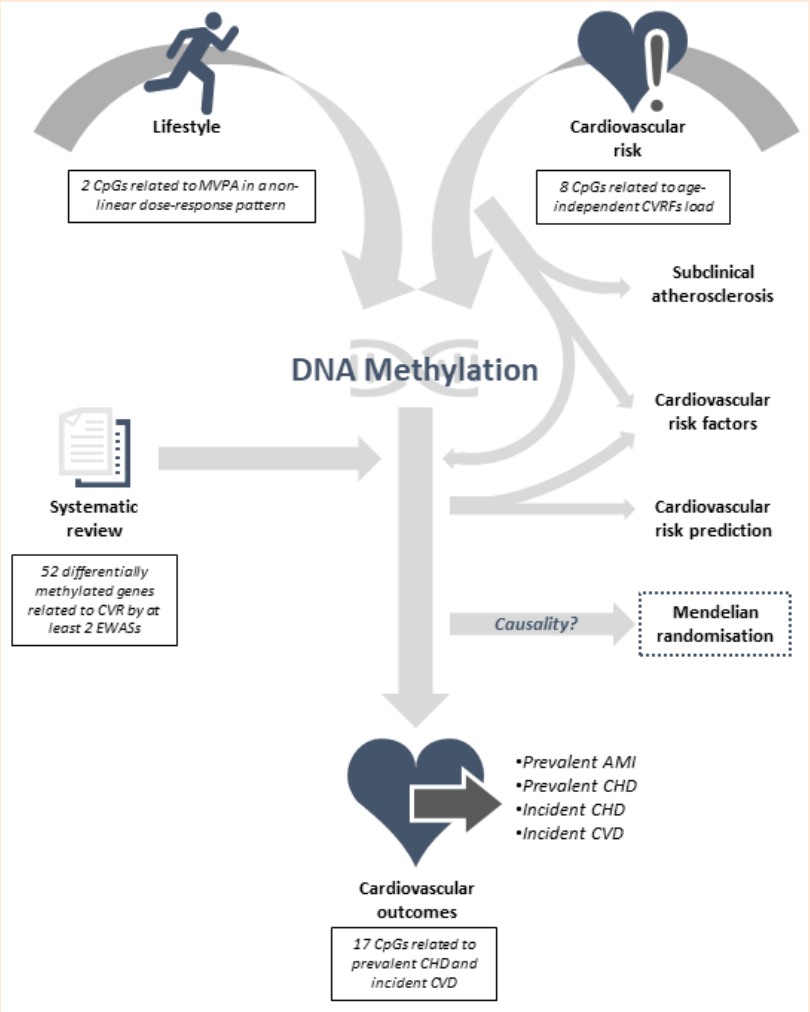
In this thesis, we have addressed the role of DNA methylation in the context of complex diseases, namely cardiovascular diseases, and some of their determinants. We analysed the association between this epigenetic mechanism and cardiovascular determinants (exposures) and clinical outcomes, and evaluated the clinical relevance of some of these findings. First, we assessed and demonstrated the relationship with a well-known lifestyle factor with multiple benefits on health, physical activity (PA). We then explored whether DNA methylomes were related to cardiovascular risk (CVR), and identified CVR-related methylation loci. Next, we moved on to study the association between DNA methylation and clinical cardiovascular outcomes, and found differentially methylated loci associated with prevalent and incident cardiovascular events, some of which were also related to CVR factors (CVRFs). In this interplay between DNA methylation and cardiovascular health, we demonstrated an association between cardiovascular events and polygenic risk scores based on some of the identified loci. However, DNA methylation loci identified neither improve the predictive capacity of a CVR function, nor were they causally associated with clinical cardiovascular outcomes (Figure 18).

To address the objectives established for this thesis, we used population-based studies based on three different cohorts: the REGICOR cohort, the Framingham Offspring Study (FOS), and the Women's Health Initiative (WHI) cohort. We performed case-control, cross-sectional, and longitudinal studies, and we applied established and new statistical methods (e.g. linear, logistic, and Cox regression; generalized additive models with smoothing splines; and surrogate variable adjustment). We also analysed DNA methylomes using standard techniques and methods; i.e. commercial arrays based on bisulfite-conversion (Infinium methylation arrays), which allowed us to test, simultaneously and at single-nucleotide level, for association with thousands of CpGs in hundreds or thousands of individuals. We used at least two populations in each analysis, and meta-analysed the results to minimize the number of false positive results (Table 2). We also combined different approaches to each objective in order to maximize the capture of informative results (e.g. PA assessed as different combinations of intensities; age-independent CVR assessed as the difference between vascular and chronological age and as the residuals of the relationship between age and estimated CVR;

association studies using hypothesis-free and hypothesis-driven approaches; prevalent and incident cardiovascular events).

Overall, this thesis supports the important role of DNA methylation in cardiovascular health, and provides new insights into the biology underlying the effects of PA and CVR on health, and the biology underlying cardiovascular events.

Figure 18. Graphical abstract of this thesis



**MVPA, Moderate-Vigorous Physical Activity; CVRF, CardioVascular Risk Factor; EWAS, Epigenome-Wide Association Study; AMI, Acute Myocardial Infarction; CHD, Coronary Heart Disease; CVD, CardioVascular Disease.*

Table 2. Summary of the main approaches and findings of this thesis.

Objective – Manuscript	Cohorts	Main analysis	Exposure	Outcome	Main result
Objective 1 – Manuscript 1	REGICOR, FOS	Two-stage EWAS	PA	DNA methylation	2 CpGs related to PA in a non-linear dose-response pattern
Objective 2 – Manuscript 2	REGICOR, FOS	Two-stage EWAS	Age-independent CVR	DNA methylation	8 CpGs related to age-independent CVR
Objective 3 – Manuscript 3	–	Systematic review	DNA methylation - CVR		52 differentially methylated genes related to CVR by at least 2 EWASs
Objective 3 – Manuscript 4	REGICOR, FOS, WHI	EWAS and candidate gene association study	DNA methylation	Prevalent AMI/CHD, Incident CHD/CVD	17 CpGs related to prevalent CHD and/or incident CVD

**FOS, Framingham Offspring Study; WHI, Women's Health Initiative; EWAS, Epigenome-Wide Association Study; PA, Physical Activity; CVR, CardioVascular Risk; AMI, Acute Myocardial Infarction; CHD, Coronary Heart Disease; CVD, CardioVascular Disease.*

4.2. Physical activity as a modulator of DNA methylation

We hypothesized that physical activity (PA) triggers changes in DNA methylation patterns. We analysed and demonstrated a relationship between PA and DNA methylation in blood cells using a hypothesis-free, two-stage cross-sectional EWAS approach performed in three independent populations. We found two CpG sites that showed differential methylation in association with PA, particularly moderate and vigorous PA (moderate- and vigorous-intensity aerobic PA, MVPA), which is the currently recommended intensity [87]. Specifically, the methylation levels of the CpGs showed a non-linear dose-response relationship with PA, mainly at high levels of MVPA.

a) Rationale and previous evidence

It is well-known that PA has multiple benefits for health [259–261], including cardio-metabolic health [85]. For the healthy adult population, current guidelines recommend at least 150 minutes of moderate-intensity aerobic PA (MPA) or 75 minutes of vigorous-intensity aerobic PA (VPA) per week, or an equivalent combination of both (MVPA) [87]. PA includes any movement that results in energy expenditure, and thus it is not limited to exercise, which is planned to improve physical fitness. At the physiological level, PA increases oxygen uptake, heart rate, blood flow, glucose metabolism, and lipolysis, as well as other responses [262]. However, its effects at the molecular level are not yet clearly understood, and the mechanisms for acute responses to exercise may be different to those for chronic adaptations to regular exercise. Previous studies suggest that different signals generated during muscle contraction result in a cascade that activates and/or represses certain pathways that regulate gene expression and protein synthesis or degradation [262]. In this context, DNA methylation and epigenetic mechanisms may be susceptible to the influence of PA. Also, DNA methylation changes related to PA may be tissue-specific.

There is some evidence of a relationship between PA and DNA methylation in different human tissues, both in a healthy and chronic disease context. Observational studies assessing the association between DNA methylation and regular PA show a weak association between these variables, although they were based on candidate gene

and global methylation studies, and had some limitations, mainly low statistical power [263,264].

Most evidence comes from interventional studies analysing different tissues, such as blood, skeletal muscle and adipose tissue. There is limited evidence of DNA methylation changes in peripheral blood cells in relation to an exercise intervention, although these changes were accompanied by transcriptional changes [265]. Very recently, Jacques *et al* performed a systematic review of studies assessing epigenetic changes (DNA methylation, miRNAs and histone modifications) in skeletal muscle following a single exercise session and a long-term training intervention (weeks to months) in healthy populations. They found that candidate-gene methylation studies mostly focused on *PGC-1a*, which is involved in regulating mitochondrial biogenesis and lipid metabolism. They also noted consistent reports that only one acute session of exercise led to dose-dependent hypomethylation of the promoter regions of that gene, accompanied by an increase in mRNA levels. The authors also concluded that, while studies that used an EWASs approach showed highly heterogeneous results, they also showed a consistent moderate effect size across the studies [266]. Regarding adipose tissue, DNA methylation changes were also reported after an exercise program, but these changes only overlapped transcriptomic changes in some genes after a long-term intervention [267,268]. These studies reported consistent hypomethylation and increased gene expression at 115 genes in adipose tissue and 64 genes in muscle. These genes various key pathways, including those related to glucose metabolism, fatty-acid synthesis, and central signalling cascades initiated by MAPK and JAK-STAT.

Simultaneously to our study (Manuscript 1), other evidence has been published that was not reviewed by Jacques *et al*. Interestingly, a study among PA-discordant adult twin pairs found no link between PA and epigenetic age based on leukocyte methylomes, although the authors acknowledged several limitations, such as small sample size [269]. Also, a study based on WGS of bisulfite-converted DNA from skeletal muscle samples from sedentary (n=8) and physically active (n=8) healthy older men found hundreds of PA-related DNA methylation-regulated genes involved in insulin sensitivity, glycolysis, oxidative stress resistance, and muscle regeneration [270]. Regarding DNA methylation changes in blood cells following acute exercise, a pilot

EWAS reported that exercise induced differences in DNA methylation of Natural Killer cells from five healthy women [271]. Finally, a prospective population-based EWAS reported several CpG sites related to PA including household chores [272]. We included this study in ours during the peer- review process, as both are population-based EWASs. However, we did not validate any of their findings, probably because there are several important differences between the two studies, such as the study design, the assessment and classification of PA, and the statistical analyses. Overall, although several epigenetic changes in response to exercise have been identified, the field of exercise epigenetics is still in its infancy, and the downstream physiological or health-related consequences require further work [266].

Finally, although PA has been shown to be associated to some clinical outcomes in a non-linear dose-response manner, previous studies have not considered non-linear associations [91,273–277]. For example, even low levels of PA are associated with reduced CHD risk, but this risk reduction tends to reach a plateau as levels of PA increase [91]. This pattern is not directly consistent with our observation of changes in DNA methylation at high levels of PA (more exactly, MVPA). Altogether, these studies highlight the importance of analysing both the linear and non-linear association between PA and clinical and biological phenotypes.

b) CpG sites related to physical activity: potential role in molecular networks

The CpG sites that showed non-linear association with MVPA were an intergenic CpG (cg24155427) and a CpG located within the gene body[21] of *DGAT1* (cg09565397). We observed that at high doses of MVPA, the higher the dose, the lower the M-values at the identified CpG sites; i.e. in the upper range of exposure, these CpGs are less methylated in individuals with higher doses of MVPA than in individuals with lower doses of MVPA. These patterns were observed in the REGICOR populations, and although they remained significant

[21] A gene body comprises introns and exons; i.e. the transcriptional region,

after meta-analysing the p-values, they were not observed in the FOS analyses. This could be due to the use of different validated questionnaires and/or to differences between the Spanish and American populations.

cg24155427 maps to an intergenic region of chromosome 1, more specifically to a CpG shelf[22] downstream of CGI chr1:31246010-31246280. This CpG site was reported to be differentially methylated in association with various immune- and inflammation-related states: smoking [278], systemic lupus erythematosus [279], aging in neutrophils of HIV patients [280], and chronic fatigue syndrome [281]. Part of the 3' sequence of *NKAIN1* is located within the CpG-upstream region up to 10,000 bp. *NKAIN1* encodes a protein that interacts with the sodium/potassium-transporting ATPase, NKAIN1 (sodium/potassium-transporting ATPase interacting 1). In a GWAS meta-analysis of five longitudinal studies in individuals of European-ancestry, a SNP in this gene was found to be associated with BMI [282]. With this information, we cannot hypothesise either about the potential role of epigenetic regulation in this genomic region, or about its functional implications. Validation studies are required to confirm that this CpG is related to PA, and to explain its role in the complex molecular networks triggered by PA and its potential clinical value.

Conversely, the information and data available for cg09565397 support the impact of PA on lipid homeostasis, with DNA methylation as a potential mediating mechanism. This CpG is located in the enhancer[23] region chr8:145,542,969-145,543,100 of *DGAT1*. One SNP close to this gene was reported to be associated with body height [283]. However, differential methylation at this CpG was not found to be associated with any trait in currently available EWASs. *DGAT1* encodes an enzyme involved in triacylglycerol synthesis – diacylglycerol O-acyltransferase 1, DGAT1 [284]. TG metabolism has been linked to PA, with regular aerobic PA decreasing TG levels in the blood [285]. Therefore, it is interesting to observe, in a population-based study of PA, a differentially methylated CpG located in a cis-

[22] A CpG shelf is a genomic region located within 2-4 kilobases up- or downstream of a CpG island.
[23] An enhancer is a regulatory region of the genome that is prone to the binding of transcription factors.

regulatory element of a gene related to TG metabolism. In addition, another mechanism may link DGAT1 levels to PA. DGAT1 has been shown to participate in the AMP-activated protein kinase (AMPK) cascade [286], a critical network for response to energetic and mitochondrial stress in eukaryotes. PA is one of the factors that leads to AMPK activation, triggered by low cellular ATP to increase ATP generation, diminish ATP consumption, and promote mitochondrial biogenesis. Some preclinical models suggest that the AMPK cascade underlies some of the benefits of PA [287]. Interestingly, exercise has been suggested to induce epigenetic regulation of *PRKAA2*, a gene encoding a catalytic subunit of AMPK [288].

At the molecular level, the 5mC at these CpGs may alter transcription factor binding, thus regulating gene expression. This may be particularly true for cg09565397, which is located within an enhancer element; regulation of transcription factor binding may in turn modulate TG levels. However, we cannot speculate about the functional consequences of any of the differentially methylated CpGs (whether they result in repression or enhancement of gene expression) because we are still at an early stage of research into the regulatory networks of epigenetic, transcriptional, and signalling factors and the intertwined relationships between the genes and their products – the so-called regulome. Importantly, one study analysing the effects of 5mCpG on the DNA binding specificities of hundreds of human transcription factors found that these effects are broader than previously reported, and that methylation can promote the binding of transcription factors in many cases [289]. In addition to the youth of regulome research, the methylation/unmethylation of these MVPA-related CpGs may have different functional outcomes in different tissues, and this is probably the case for cg09565397 because enhancers are usually tissue-specific [290].

c) Clinical relevance of the identified CpG sites

Overall, our findings suggest that the recommended dose of PA and MVPA produce changes related to differential DNA methylation in the general adult population. The translational potential of PA-related epigenetic markers could allow us to establish personalized exercise routines that modulate the epigenome and, in turn, contribute to the prevention of several chronic diseases, or to control their progression

[264,291]. However, currently there is only scarce evidence about the complex biological network that is triggered by PA and results in a functional outcome; the following limiting issues should be considered. As stated by Rönn and Lind in an editorial on this topic, for future studies "to get a picture of the overall regulatory effect of exercise on the epigenetic profile in human tissues and on metabolism, we need to consider several different tissues at the same time point and try to understand how they interact" [292]. Moreover, we need to consider a more complete picture of the epigenome, including and integrating DNA methylation, non-coding RNA, histone modifications, and transcriptome.

In this regard, although certain PA intensities or regimes are related to DNA methylation changes, we cannot infer whether these changes are beneficial for health in themselves, or whether they simply mediate the beneficial effects of PA on health. We also need to take into account the fact that different individuals respond differently to the same dose of PA [293], and this type of biomarkers could shed some light on the mechanisms explaining this inter-individual variability.

Highlights of Objective 1

- PA was associated with blood DNA methylation in a two-stage cross-sectional EWAS
- Two differentially methylated sites were associated with MVPA in a non-linear dose-response way
- One of the CpGs suggests that TG metabolism is involved in the mechanisms triggered by PA

4.3. DNA methylation and cardiovascular risk factors load

Since CVRFs are known to be individually related to DNA methylation, we hypothesized that the accumulation of CVRFs may have a cumulative effect on the methylome. However, when considering accumulated CVRFs, we excluded age because of its impact on susceptibility to complex diseases, and because of the close relationship between epigenetic mechanisms and aging. For this purpose, we defined two approaches, one using the difference between vascular and chronological age, and another using the residuals of the linear association between estimated CVR and age. Using a two-stage cross-sectional EWAS design, we found an association between age-independent estimated CVR and the blood methylome in individuals from two independent populations. We found eight differentially methylated sites that were associated with age-independent CVR, all of which were related to individual classical CVRFs, and all but one to at least smoking. We integrated these sites into polygenic risk scores, and assessed whether these scores were predictive of subclinical atherosclerosis and incident cardiovascular and coronary events.

a) Rationale and previous evidence

Classical CVRFs are widely known to trigger cardiovascular events as well as subclinical intermediate phenotypes. The cumulative and synergistic effect of CVRFs on cardiovascular health is well documented [294], as is the fact that CVR increases throughout life [26] (see sections 1.1.a and 1.2.a). Risk functions usually estimate the global effect on CVR of the accumulation of classical CVRFs as an absolute CVR (probability of having a clinical cardiovascular event, usually in the following 10 years; section 1.3.a). However, the entire picture of the molecular mechanisms and networks that mediate the crosstalk between CVRFs and physiological outcomes has not been completely developed. Since most classical CVRFs are individually associated with DNA methylation patterns, the accumulation of CVRFs is also very likely to be related to specific DNA methylation patterns. Both CVR and DNA methylation are highly influenced by aging, although age-independent CVR resulting from the confluence of other CVRFs may help to decipher molecular mechanisms linking

DNA methylation and CVR other than those that are directly related to aging.

A growing number of EWASs focused on classical CVRFs have been published in the past decade, and we now have broad evidence on DNA methylation loci that are related to CVRFs. For instance, in the REGICOR study, a recent doctoral thesis addressed the association between DNA methylation and some individual classical CVRFs, including smoking, lipid levels, and obesity [295]. Regarding smoking, using the REGICOR cross-sectional study sample as a discovery population, Sayols-Baixeras found a reversible exposure-time-dependent relationship between smoking and DNA methylation, and confirmed several methylation loci that had previously been reported to be related to smoking, such as cg05575921 in *AHRR*, which we also found in the second and third objective of this thesis (see below). They also identified one additional novel CpG [296]. Since the publication of these findings, other EWASs on smoking have been performed, and the EWAS databases now contain thousands of entries for "smoking" (7,654 entries in the EWAS catalog, 14,273 sites in the EWAS Atlas), reported by 46 publications according to a search performed on 27th August 2019. Concerning serum lipid levels, a two-stage EWAS in the REGICOR cross-sectional study sample (discovery) and the FOS sample (validation) confirmed previously reported methylation loci related to lipids, as well as six new gene-mapping sites [297]. In the EWAS catalog there are 698 entries for "lipids", while the EWAS Atlas contains 10 sites and 5 publications for "blood triglyceride levels", one site and one publication for "blood LDL levels", one site and one publication for "total cholesterol", and seven sites and three studies for "blood HDL levels". As a final example, and from the REGICOR study, a two-stage EWAS in the cross-sectional study sample (REGICOR and FOS) identified 94 CpGs that were related to both obesity and BMI, 70 of which were novel findings [298]. In the EWAS catalog there are 7 and 1,211 entries for "obesity" and "BMI", respectively, while in the EWAS Atlas, there are 8,134 and 1,797 sites, and 7 and 17 publications, respectively.

At the time of writing this thesis, no study had assessed the association between vascular age and DNA methylation. Interestingly, during publication of the article addressing the second objective of this thesis, another study by the Atherosclerosis Risk in Communities (ARIC)

consortium reported that epigenetic aging of blood cells (mentioned in section 1.4.e) is a predictor of both subclinical atherosclerosis (assessed as carotid intima-media thickness) and incident cardiovascular events, independently of chronological age and classical CVRFs [299]. This finding supports the notion that age-independent CVR is related to DNA methylation, and that other variables different from individual classical CVRFs could potentially be used as CVR markers. In addition, DNA methylation at specific sites has also been analysed in association with subclinical atherosclerosis assessed as intima-media thickness or pulse wave velocity [300–302]. Our findings suggest that the synergistic effects of CVRFs (excluding chronological age) influence DNA methylation and that the resulting marks could potentially be combined as markers of both subclinical atherosclerosis and of the risk of incident cardiovascular events.

b) CpG sites associated with age-independent cardiovascular risk: potential role in molecular networks

We identified eight age-independent CVR-related CpGs that, when combined in polygenic risk scores, are predictive of incident CVD. Four of these are intergenic (cg12547807, cg27537125, cg05951221, and cg21566642) and are related to smoking, consistent with previous findings [278,303]. cg12547807 is also associated with BMI. The first two CpGs map to regulatory regions of chromosome 1: cg12547807 to the enhancer at chr1: 9,473,003-9,474,074 and cg27537125 to the promoter-associated DMR region chr1:25,348,676-25,349,815. The last two CpGs are located in the CGI chr2:233,283,397-233,285,959, and are thus considered to be the same locus for the purpose of constructing the risk scores. The genomic location of these four CpGs indicates that it would be interesting for future studies to also analyse DMRs.

The other four CpGs (cg19939077, cg18608055, cg05575921 and cg00574958) map to gene sequences (the gene bodies of *PPIF, SBNO2,* and *AHRR,* and the 5'UTR[24] of *CPT1A,* respectively):

- ***PPIF*** encodes peptidylprolyl isomerase F (also known as cyclophilin D), which is involved in ischemia/reperfusion injury, heart failure, arterial thrombosis, cardiac hypertrophy, atherosclerosis, and diabetes [304]. cg19939077 maps to the promoter-associated regulatory region chr10:81,106,660-81,108,439 located within the upstream shore[25] of the CGI chr10:81107082-81107488. Previous EWASs reported it to be associated with smoking [278] (validated in our study) and alcoholism [305,306]. We also found it to be associated with BMI.

- ***SBNO2*** expresses a component of the pathways leading to the anti-inflammatory effect of IL-10 [307]. cg18608055 maps to the promoter-associated regulatory DMR chr19:1,130,697-1,131,291 located within the upstream shore of the CGI chr10:81107082-81107488. It was previously found to be associated with smoking [278] and BMI [308,309] (consistently with our study), as well as with C-reactive protein [310] and the cardiovascular biomarker GDF-15 [311]. In this last study it was also non-significantly hypomethylated in individuals that had suffered a MI, which is consistent with our findings in relation to age-independent CVR.

- ***AHRR*** is the most epigenetically regulated smoking-related gene [251]. It encodes the aryl-hydrocarbon receptor (AhR) repressor (AhRR), which is involved in repressing the AhR signalling cascade (activated when bound by toxins from cigarette smoke) [312] and other signalling pathways (e.g. NF-$\varkappa$B). AHRR was also found to be a tumour suppressor and a modulator of inflammatory responses [313]. Its target, AhR, is a transcription factor that integrates environmental, microbial, metabolic and endogenous signals to control adaptation to the cellular environment. It

[24] The 5'UTR is the untranslated region located upstream of a protein-coding sequence on a transcript, which encompasses a sequence recognized by the ribosome for translation initiation.

[25] A CpG shore is a genomic region located within 2 kylobases up- or downstream of a CpG island.

modulates several biological processes in complex contexts that are relevant for pathological conditions (e.g. inflammatory, autoimmune, neoplastic, metabolic, and degenerative diseases) [314]. cg05575921 maps to an enhancer located in the downstream shore of the CGI chr5:373,842-374,426. Its methylation status was shown to be linked to smoking status, and had the strongest association in most EWASs assessing blood methylation in relation to smoking [278,296,303]. This CpG also accurately discriminates between current and never smokers [315,316], and was associated with prenatal exposure to maternal smoking [317], high smoking-related morbidity (such as lung cancer) [318–320] and all-cause mortality [318,321].

- ***CPT1A*** encodes carnitine palmitoyltransferase 1A, which is critical for allowing long-chain fatty acids to enter into the mitochondria for subsequent oxidation [322]. cg00574958 lies close to SNP rs78442314, and maps to the downstream shore of the CGI chr11:68,608,155-68,609,419 and to CGI chr11:68,364,198-68,364,338 (the latter predicted by hidden Markov models)[26]. Consistently with our findings, this CpG was previously associated with metabolic syndrome [323], diabetes [324,325], BMI [309,326,327], and lipid levels [328–331].

We found that all eight of these CpGs were hypomethylated in association with age-independent CVR. Several studies suggest that smoking results in upregulated *AHRR* expression [332–338], which can also be inferred from our results, as the CpG we found in this gene is highly associated with smoking but is not related to any other CVRF. The lack of the methyl group at this site, which maps to an enhancer, may promote the binding of transcription factors that enhance gene expression.

Conversely, as explained in section 4.2.b, with the currently available data and literature, we cannot anticipate the biological outcomes of

[26] While the other reported CGIs are those in the UCSC Genome Browser and derived using algorithms based on the definition of CGI proposed by Gardiner-Garden and Frommer in the late 1980s [379], these CGIs are based on a non-human-exclusive and more modern definition and predicted with probability scores constructed using hidden Markov models [380].

any of the other seven CpGs because of the complexity and tissue-specificity of the regulome. Interestingly, four of the CpGs map to regulatory regions, two are part of the same CGI, and the other lies in a highly CpG-enriched region (the shore of a CGI). This suggests that these regions are important for controlling gene expression in the presence of a combination of CVRFs. Moreover, the genes that are subject to this epigenetic regulation are related to processes or pathways that are important for the development of atherosclerosis and CVD such as inflammation, fatty acid catabolism, and multifunctional pathways.

c) Clinical relevance of the identified CpG sites

Our findings suggest that the confluence of several CVRFs is related to differential DNA methylation independently of age in the general adult population. Importantly, we provide new evidence indicating that polygenic risk scores including epigenetic marks related to CVRF load are predictive of clinical cardiovascular events. While the identified epigenetic markers are not as clearly predictive of subclinical atherosclerosis, they still highlight the potential of further exploring the interaction between CVR-related DNA methylation and markers of atherosclerosis. Another important issue is that, while most of the DNA methylation markers identified were related to smoking, their association with CVD remained significant even after adjusting for smoking. Thus, DNA methylation may more accurately reflect smoking exposure than standard questionnaires.

Our point in the section 4.2.c about the clinical relevance of PA-related methylation markers may also be extrapolated to that of the markers associated with age-independent CVR. The translational potential of CVR-related epigenetic markers may allow clinicians and epidemiologists to predict the CVR of individuals more accurately. Thus, they may also improve clinical decision making about the management of CVRFs and intermediate cardio-metabolic phenotypes, and could also be used as research tools to explore atherosclerosis pathogenesis. Although DNA methylation from whole blood cells is a fairly accurate proxy for that in other tissues [317,334,339], it would be interesting to study, at the same time point, the methylome of other tissues that commonly altered in the presence of CVRFs. In addition, other layers of information shaping the

regulome should be considered, and the functional consequences of those changes should be explored.

It is necessary to further assess their potential as targets for new therapies, and to determine the causality of the observed associations. Our results indicate that the direction of these associations may go from the load and synergistic effects of CVRFs to the clinical outcomes, with DNA methylation as one of the potential mediating mechanisms. This hypothesis is more plausible smoking, which triggers epigenetic modifications, but the scenario where all the CVRFs act synergistically may be more complex, and the direction of the association between DNA methylation and the CVR may vary depending on the CVRF or the methylation loci. We cannot rule out the possibility that specific DNA methylation patterns can influence either the presence of some CVRFs (e.g. biochemical traits) or their effects on downstream molecular and physiological pathways. This possibility was described for TG levels, precisely for the CpG in *CPT1A* that we also found to be related to age-independent CVR and TG levels [329,330]. In the MR study by Sayols-Baixeras *et al* both directions of the association were confirmed.

Highlights of Objective 2

- Age-independent CVRF load was associated with blood DNA methylation in a two-stage cross-sectional EWAS
- Eight differentially methylated sites were found, all of which were related to individual classical CVRFs
- Their genomic location suggests that they may control the expression of genes related to key atherosclerosis processes, such as inflammation, fatty acid catabolism, and multifunctional pathways
- Genetic risk scores including these CpGs were found to be associated with incident CVD, independently of classical CVRFs

4.4. DNA methylation and cardiovascular disease

For the last objective of this thesis, we studied the relationship between DNA methylation and cardiovascular outcomes. We first performed a systematic review of the topic, and then explored the association between DNA methylation and clinical cardiovascular outcomes, using both EWAS and hypothesis-driven approaches based on data from three independent populations. The latter approach encompassed the loci found to be differentially methylated in association with age-independent CVR (Manuscript 2), as well as the genes identified by the systematic review to be consistently differentially methylated in relation to atherosclerosis. In both the EWAS and candidate gene approaches, we found differentially methylated loci associated with prevalent and incident cardiovascular outcomes, some of which were also related to classical CVRFs. We confirmed that polygenic risk scores based on loci that are differentially methylated in association with incident CVD were also related to risk of cardiovascular events. However, these scores did not improve the predictive capacity of the classical risk functions, and we could not determine whether the CpGs were causally associated with CHD.

a) Rationale and previous evidence

Different approaches have been used to study the association between DNA methylation and CVD in distinct tissues and populations, and using distinct study designs. Revising published evidence on the relationship between DNA methylation and atherosclerotic outcomes, we found that most studies that matched our eligibility criteria assessed DNA methylation using candidate gene association designs, some reporting consistent associations for some genes (*ESRa*, *ABCG1*, *FOXP3*, *IL-6)*. Studies assessing global methylation levels showed inconsistent results, while those that interrogated the whole methylome using Illumina arrays identified 84 genes as consistently differentially methylated (52 in the same direction: hyper- or hypomethylated). One-third of these genes had been linked to obesity by GWAS, and *in silico* functional analysis identified several diseases and functions linked to inflammation, metabolism and cardiovascular disease. Limited work has been done

on the association between DNA methylation and incident CVD, either using only global methylation levels [340], or reported DNA methylation markers that provided only moderately improved predictive value over classical CVRFs [341].

More than two years after the online search performed for the systematic review, the number of articles returned using the same search terms ["DNA methylation" AND ("Coronary heart disease" OR "Ischemic heart disease" OR "Myocardial infarction" OR "Cardiovascular risk" OR "Vascular age")] has increased from 96 to 163 (August 29th, 2019), including some notable ones. Three interesting non-prospective EWASs have been reported, although they have some limitations. Nakatochi *et al* performed a case-control EWAS to examine the association between blood methylomes and MI in a population of elder Japanese men. They found two CpGs in genes containing MI-related SNPs although the CpGs were independent of these SNPs [342]. Wang *et al* conducted a cross-sectional EWAS comparing (intra-individually) different vascular tissues in six patients with CHD who underwent coronary artery bypass surgery, and found loci that are enriched in pathways related to immune responses and metabolism [343]. In cardiac samples from 11 individuals with heart failure (HF), Pepin *et al* reported that DNA methylation is a key mechanism in the metabolic reprogramming that occurs in ischemic HF, providing biomarkers that distinguish ischemic HF from other HF aetiologies of [344].

In addition, non-prospective NGS studies analysing blood methylomes have also been performed using RRBS. In one study carried out in patients with heart failure and age- and sex-matched controls, Li *et al* identified three differentially methylated genes [345]. More recently, one NGS-based analysis in 32 young MI patients found differential methylation profiles between MI patients with and without recurrent events within one year of the first event. The loci they identified are involved in pathways such as cardiac function, and repair and response to injury. Interestingly, individuals with recurrent cardiac events showed hypermethylation at *DGAT1* (a gene that we found in the objective addressing PA and DNA methylation) [346].

Regarding prospective studies, some associations between DNA methylation and CHD have been discovered in the second step of

studies exploring its association with CVRFs [347,348]. Also, in the ARIC study epigenetic age was predictive of fatal CHD, peripheral artery disease, and heart failure, independently of chronological age and other traditional CVRFs [299]. More importantly, simultaneously to our study, some differential methylation studies of incident CVD as the primary outcome have been performed using different analytical strategies, and have provided different findings [349–351]. Westerman *et al* analysed DMRs and epigenetic modules in association with CVD in two of the samples we also studied (FOS and WHI), and identified mechanisms related to development and monocyte biology [349][27]. Currently, they are aggregating the results into a direct predictor of CVR, which is being validated in the REGICOR case-control study in collaboration with us (unpublished data). Ward-Caviness *et al* provided more insights into the association between DNA methylation and incident MI using three different cohorts. They found nine CpGs that were further shown to be related to the metabolism of some amino acids, and that are moderately predictive of MI events [350]. Finally, Agha *et al* recently published a powerful multi-centre meta-analysis of leukocyte- and 450k-based EWASs of incident CHD in nine population-based cohorts from the USA and Europe (sample size >11,000 individuals). They reported 52 CpGs related to incident CHD that mapped to regulatory regions of lncRNA regions, and provided evidence supporting the causality of the associations [351]. Taken together, these and our findings (summarized in Table 3) highlight the complexity of the epigenetic mechanisms involved in the intricate networks underlying CHD, and CVD in general. They also highlight the challenge of establishing a consensus on the best analytical approach for comparing and integrating data from different studies.

Table 3. Summary of CpGs related to CHD in EWASs.

CpG	Chr*	Position (bp)	Gene	Study
cg21609024	1	53,795,111	*LRP8*	[350]

Continued

[27] Note that this work is still under the revision for publication in a peer-reviewed journal.

cg10073091	1	55,352,301	*DHCR24*	[350]
cg11955541	1	145,040,160	*PDE4DIP*	[350]
cg03458344	1	170,964,477	*C1orf129*	[350]
cg00699486	6	166,144,768	Intergenic	[350]
cg07311024	12	75,785,089	*GLIPR1L2*	[350]
cg08193363	13	32,605,254	*FRY*	[350]
cg23074119	14	78,174,751	*ALKBH1; C14orf156*	[350]
cg23541257	19	18,096,662	*KCNN1*	[350]
cg12766383	1	19,403,306	*UBR4*	[351]
cg07475527	1	24,864,545	Intergenic	[351]
cg10702366	1	60,070,383	*FGGY*	[351]
cg07015775	1	86,174,125	*ZNHIT6*	[351]
cg00466121	1	86,174,151	*ZNHIT6*	[351]
cg00355799	1	244,109,212	Intergenic	[351]
cg23245316	2	3,260,005	*TSSC1*	[351]
cg03467256	2	10,556,515	*HPCAL1*	[351]
cg13822123	2	54,197,256	*PSME4*	[351]
cg26470101	2	173,099,597	Intergenic	[351]
cg06639874	2	238,417,703	*MLPH*	[351]
cg14029912	3	5,027,616	Intergenic	[351]
cg22617878	3	10,417,183	*ATP2B2*	[351]
cg07289306	3	44,039,357	Intergenic	[351]
cg06582394	3	121,902,622	*CASR*	[351]
cg00393373	4	10,456,597	*ZNF518B*	[351]
cg08853494	4	76,439,657	*RCHY1; THAP6*	[351]
cg02155262	4	178,363,707	*AGA*	[351]
cg02683350	5	178,658,501	*ADAMTS2*	[351]
cg19935845	6	32,074,856	*TNXB*	[351]
cg14010194	6	42,152,817	*GUCA1B*	[351]

Continued

cg21018156	6	134,061,814	Intergenic	[351]
cg05892484	7	2,143,507	*MAD1L1*	[351]
cg24977276	7	74,105,270	*GTF2I*	[351]
cg24423782	7	129,410,417	*MIR182*	[351]
cg02321112	7	156,810,523	Intergenic	[351]
cg25497530	7	158,059,944	*PTPRN2*	[351]
cg26042024	8	135,610,009	*ZFAT; ZFATAS*	[351]
cg05820312	8	141,468,672	*TRAPPC9*	[351]
cg14185717	9	16,864,746	*BNC2*	[351]
cg13827209	9	101,912,842	*TGFBR1*	[351]
cg19227382	10	73,521,606	*CDH23; C10orf54*	[351]
cg07436807	10	90,712,767	*ACTA2*	[351]
cg10307345	11	18,771,567	*PTPN5*	[351]
cg17556588	11	32,854,320	*PRRG4*	[351]
cg24318598	11	70,034,186	*ANO1*	[351]
cg03031868	13	47,371,523	*ESD*	[351]
cg18598861	14	24,635,529	*IRF9*	[351]
cg20000562	14	36,978,633	*SFTA3*	[351]
cg04987302	14	57,476,116	Intergenic	[351]
cg22871797	15	22,992,526	*CYFIP1*	[351]
cg25196881	15	39,780,412	Intergenic	[351]
cg26467725	15	92,647,041	*SLCO3A1*	[351]
cg06596307	15	99,405,016	*IGF1R*	[351]
cg24447788	19	795,310	Intergenic	[351]
cg01751802	19	11,309,639	*KANK2*	[351]
cg09777776	19	24,269,890	*ZNF254*	[351]
cg06442192	19	48,059,856	*ZNF541*	[351]
cg02449373	19	49,256,123	*FUT1*	[351]
cg22618720	20	55,959,549	Intergenic	[351]

Continued

cg08422803	21	46,341,067	*ITGB2*	[351]
cg20545941	22	43,821,227	*MPPED1*	[351]
cg25103337	1	9,293,583	*H6PD*	Manuscript 4
cg19893751†	2	45,029,285	Intergenic	Manuscript 4
cg21566642	2	233,284,661	Intergenic	Manuscript 4
cg11643285	3	16,411,667	*RFTN1*	Manuscript 4
cg17238319	3	16,428,391	*RFTN1*	Manuscript 4
cg08122652	3	122,281,939	*PARP9; DTX3L*	Manuscript 4
cg00076653	4	15,341,878	*C1QTNF7*	Manuscript 4
cg05575921	5	373,378	*AHRR*	Manuscript 4
cg21429551	7	30,635,762	*GARS*	Manuscript 4
cg19390658	7	30,636,176	*GARS*	Manuscript 4
cg09165129†	7	137,524,369	*DGKI*	Manuscript 4
cg00574958	11	68,607,622	*CPT1A*	Manuscript 4
cg19314882	11	117,391,953	*DSCAML1*	Manuscript 4
cg14597545	15	73,074,210	*ADPGK*	Manuscript 4
cg03636183	19	17,000,585	*F2RL3*	Manuscript 4
cg07817505†	19	17,972,324	*RPL18AP3; SNORA68; RPL18A*	Manuscript 4
cg06500161	21	43,656,587	*ABCG1*	Manuscript 4

**Chr, chromosome.* †*CpG must be considered with caution (see incident CHD).*

b) DNA methylation and cardiovascular outcomes in incident and prevalent cases

DNA methylation patterns associated with cardiovascular outcomes may differ between individuals with prevalent and incident disease. DNA methylation changes may occur before the event, either as a driving factor or as a result of other changes. However, other methylome changes could be caused by the response to the acute event, or to the medical treatment received. Therefore, we reasoned

that analysing both sides of the path could allow us to better understand cardiovascular epigenetics in general, with implications for the improved clinical risk assessment. By using EWAS and candidate gene strategies, we have identified novel methylation loci associated with prevalent AMI and CHD, as well as new loci related to incident CHD and CVD, some of which are differentially methylated in association with both prevalence and incidence. Additionally, not all of them were further shown to be associated with classical CVRFs, suggesting that other related factors and pathways may be important in the development of CHD and CVD.

- **CpG sites related to both prevalent and incident CVD**

Three CpGs that we also found to be differentially methylated in association with age-independent CVR (intergenic cg21566642, the *AHRR*-regulator cg05575921, and the *CPT1A*-regulator cg00574958) were consistently linked to prevalent AMI and CHD, and to incident CVD. Also, three other CpGs were consistently related to both prevalent and incident events: cg06500161, cg21429551 and cg19390658, which are located in *ABCG1* and *GARS*. *ABCG1* encodes a transporter involved in lipid homeostasis [352,353]. *GARS* encodes the glycyl-tRNA synthetase, and is thus critical for translation; variants in this gene have also been found to be related to mitochondrial dysfunction [354].

All six CpGs were further related to classical CVRFs, and highlight smoking and lipid metabolism as determining factors. Altogether, the findings for these CpGs suggest that they reflect specific regulomic mechanisms where gene expression mediates the synergistic effect of CVRFs on cardiovascular health, resulting in acute events. This intermediary role may occur at any age in the case of the three CpGs that we also found in objective 2 of this thesis. The fact that these CpGs were not differentially methylated in association with incident CHD could be due to differences between the WHI and the FOS populations or the low number of CHD events in FOS.

- **CpG sites related to prevalent CHD**

Four additional CpGs were related to prevalent CHD and also to classical CVRFs: cg03636183, cg00076653, cg11643285 and cg17238319, which map to *F2RL3*, *C1QTNF7* and *RFTN1*. Finally, other two CpGs were associated with prevalent CHD but not with CVRFs: cg14597545 and cg19314882, which map to *ADPGK* and *DSCAML1*, respectively.

- ***F2RL3*** encodes a protease-activated receptor that is key for platelet activation and cell signalling. The CpG we identified is widely reported to be associated with smoking [251,296], as we also found in the analysis of CVRFs. It has also been linked to all-cause and cardiovascular mortality [253,355,356].

- ***C1QTNF7*** encodes a member of the CTRP family of adipokines related to metabolic dysfunction in the context of obesity [357].

- ***RFTN1*** encodes a protein of lipid rafts, which are specialized cell-membrane microdomains that are enriched in signalling molecules. It is involved in regulating platelet signalling by adding long chain fatty acids to the protein, a reversible modification known as palmitoylation [358].

- ***ADPGK*** encodes a glucokinase that catalyses the ADP-dependent phosphorylation of glucose in the first step of glycolysis. Its role is relevant during ischemia/hypoxia, as it would save ATP, the primary source of energy in glucose catabolism [359].

- ***DSCAML1*** encodes a cell surface adhesion protein that is involved in neuronal differentiation. A recent study found that a SNP in this gene is related to glycated serum protein, a measure of glycaemia [360].

The fact that these six CpGs were not related to incident events suggests that they may not cause the CHD event, but rather are a consequence of it. This may be especially true for the two CpGs that were not related to any CVRF (regulating *ADPGK* and *DSCAML1* expression). However, the CpGs that were further associated with CVRFs question this hypothesis, and thus highlight the complexity

of the molecular network underlying physiological functions that affect cardiovascular outcomes.

- **CpG sites related to incident CHD and CVD**

The CpGs found to be related to incident CHD were not associated with CVRFs: cg19893751, cg09165129 and cg07817505. In the EWAS on incident CHD, we could not correct the genomic inflation[28] of the results in the WHI sample, so these three CpGs must be considered with caution and must be validated in further analyses.

Apart from the CpGs already described, other two CpGs were related to incident CVD: cg08122652 and cg25103337, which are located within *PARP9/DTX3L* and *H6PD*, respectively.

- ***PARP9/DTX3L*** is involved in immune response [361], and participates in the pro-inflammatory IFN-γ signalling that controls macrophage activation [362]. Interestingly, the CpG we found has recently been linked to lower risk of incident CHD [348].

- ***H6PD*** encodes one enzyme involved in the pentose phosphate pathway, which is critical for glucose metabolism. Over-activation of this pathway was shown to trigger vascular inflammation linked to hyperglycaemia [363]. Deficiency of the enzyme was also found to be inversely associated with atherosclerosis and CHD [364,365].

Taken together, the findings on the association with incident events suggest that those methylation signatures may be causally related to CVD. The loci that are not related to traditional CVRFs also support the relevance of inflammation in the development of CHD.

[28] Statistical inflation is an important concern in association analyses that perform multiple comparisons. Genomic inflation is reported by calculating the genomic inflation factor lambda, and can be visualized using quantile-quantile (QQ) plots. The higher the inflation is, the higher the false positive rate.

Table 4 summarizes the genomic location and features of the 17 CpGs. Interestingly, 15 of them map to gene sequences, mostly the gene body, while two map to the 5'UTR and one to the promoter. Five are located within enhancer elements, and three within regulatory regions associated with promoters (one cell-specific). Overall, these observations suggest that these regions control the expression of these genes in the context of cardiovascular events, some before the event, and others as a consequence of the event. Furthermore, the regulated genes are related to processes that are important for CHD and CVD, such as inflammation. However, as explained in section 4.2.b and 4.3.b, due to the complexity and tissue-specificity of the regulome we cannot predict the biological outcome of the differential methylation status of the cardiovascular-related CpGs based only their genomic features and product function.

Table 4. Genomic features of the 17 CpGs related to CHD and/or CVD.

CpG	Chr*	Position (bp)	Gene	Part of the gene	CGI*	Regulatory element	Probe SNPs
cg25103337	1	9,293,583	*H6PD*	TSS1500	N_Shore; chr1: 9,294,401-5,118	–	–
cg19893751	2	45,029,285	Intergenic	–	chr2: 45,028,955-9,318	–	–
cg21566642	2	233,284,661	Intergenic	–	chr2:233,283,397-5,959	DMR	–
cg11643285	3	16,411,667	*RFTN1*	Body	–	Enhancer	rs61745975; within 10bp
cg17238319	3	16,428,391	*RFTN1*	Body	–	Enhancer	rs712875; 10-50 bp
cg08122652	3	122,281,939	*PARP9; DTX3L*	5'UTR; TSS1500	N_Shore; chr3: 122,283,002-594	–	–
cg00076653	4	15,341,878	*C1QTNF7*	Body	–	–	–
cg05575921	5	373,378	*AHRR*	Body	N_Shore; chr5: 373,842-4,426	Enhancer	–

Continued

cg21429551	7	30,635,762	*GARS*	Body	S_Shore; chr7: 30,634,124-5,058	Promoter Associated; chr7:30,635,729-842	–
cg19390658	7	30,636,176	*GARS*	Body	S_Shore; chr7: 30,634,124-5,058	Promoter Associated; chr7:30,636,081-376	–
cg09165129	7	137,524,369	*DGKI*	Body	–	Enhancer	–
cg00574958	11	68,607,622	*CPT1A*	5'UTR	N_Shore; chr11: 68,608,155-9,419	–	rs78442314; 10-50 bp
cg19314882	11	117,391,953	*DSCAML1*	Body	S_Shelf; chr11: 117,389,186-471	–	–
cg14597545	15	73,074,210	*ADPGK*	Body	N_Shore; chr15: 73,075,586-6,663	Promoter Associated and Cell type specific; chr15:73,074,191-372	–
cg03636183	19	17,000,585	*F2RL3*	Body	N_Shore; chr19: 17,000,627-1,398	DNase I Hyper-sensitivity Site	–
cg07817505	19	17,972,324	*RPL18AP3; SNORA68; RPL18A*	Body; TSS1500; Body	S_Shore; chr19: 17,970,454-1,123	–	–
cg06500161	21	43,656,587	*ABCG1*	Body	S_Shore; chr21: 43,654,846-5,465	Enhancer	rs9982016; within 10bp

**Chr, chromosome; CGI, CpG island.*

c) Clinical relevance of the identified CpG sites

DNA methylation biomarkers related to cardiovascular outcomes have great potential for use in preventive and clinical medicine. For instance, an assay has recently been developed for detecting cardiomyocyte death (which occurs in MI events or in abnormal cardiac physiology and development) based on methylation levels in cardiomyocytes. These methylation biomarkers enable to quantify cardiomyocyte DNA in circulating cell-free DNA derived from dying cells [366]. Despite recent advances in cardiovascular epigenetics, more studies with novel methodologies are required to identify powerful methylation biomarkers.

• CpGs as potential predictive biomarkers of CVD

The CpGs found to be associated with cardiovascular events are expected to provide complementary information to that offered by classical CVRFs, mainly those that are associated with incident events. Thus, the CpGs identified could be used as predictive markers of future CV events. They could be included in CVR functions to evaluate whether they add predictive value over that provided by CVRFs. However, considering their effect sizes, a more powerful approach would be to integrate them into risk scores. Thus, only one variable would be added to the CVR function but would capture all the epigenetic marks linked to this type of event.

We adhered to AHA and European recommendations on assessing the value of risk scores as CVR biomarkers [128,129]. The risk scores based on CpGs that are differentially methylated in relation to incident CVD were also predictive of 10-year risk of an event, independently of the classical CVRFs. However, they did not improve the capacity of the Framingham risk function to discriminate events, nor did they improve the reclassification of the individuals into risk categories. This could be due to uncomplete capture of the variability, their dynamic performance throughout life, or to their redundant encrypted information. Whatever the reason for this lack of improvement, it is not surprising to find biomarkers that do not improve CVR functions. For instance, in the REGICOR study, the inclusion of SNP-based risk scores in a CVR function did not improve the discrimination capacity of classical risk functions. However, Lluíis-Ganella *et al* showed a more accurate reclassification when including risk scores that capture

genetic variants, and the reclassification was especially improved in the intermediate risk category [367]. In this case, the markers were constant throughout life.

In the first study assessing DNA methylation and MI (a two-stage case-control EWAS) [341], the authors also evaluated the improve performance of MI risk prediction when including the differentially methylated region within *ZBTB12* and the LINE-1 hypomethylation patterns they found. Interestingly, they reported improved discrimination between MI cases and controls and better prediction accuracy (discrimination improved but not significantly). The study design was not prospective, the sample size was smaller, and discrimination was assessed by comparing the Area Under the Receiver Operating Characteristic Curves; more importantly, instead of creating risk scores, they clustered individuals into four classes using an algorithm based on the methylation of the DMR in *ZBTB12* plus a fifth class based on LINE-1 methylation. The five classes were included in the new MI risk-prediction model. These methodological differences could explain why the clinical utility of methylation biomarkers reported by this study is not consistent with ours. More recently, Ward-Caviness *et al* developed an epigenetic fingerprint based on the nine CpGs they identified, and showed that it moderately discriminated incident MI cases, although they did not evaluate whether this lead to an improvement over classical CVRFs [350]. Westerman *et al* have developed risk scores based on the methylation profiles they found in their previous study [349], and their potential as predictors is being assessed using new complex methods that consider heterogeneity across populations (unpublished data). In addition, epigenetic-based aging metrics have been shown to predict CVD, although they do not provide clinically meaningful improvements in discrimination [368,369].

Compared to other work on CVR estimation using DNA methylation markers, overall our findings show that CVR prediction and the discovery of novel predictive biomarkers remains challenging. CVR due to DNA methylation could either be redundant, or could provide complementary information to existing risk metrics, such as genetic scores [191] or those based on classical CVRFs. A better approach would probably be to integrate different layers of biological information to predict CHD and CVD. For instance, a recent study using machine learning techniques in the Framingham population

showed the potential of using an integrated genetic-epigenetic-phenotypic approach as an alternative to the conventional one based only on CVRFs [257].

- **CpGs as potential therapeutic targets and prognostic biomarkers of CVD**

Deciphering the causality of the associations identified in objective 3 of this thesis would have allowed us to identify novel targets for therapeutic studies, as well as prognostic markers that are useful for the choice of clinical intervention. However, we only had access to one public meQTLs database, the Accessible Resource for Integrated Epigenomic Studies (ARIES) [370], but not genotype data for any of the three populations included in this thesis. Studies analysing the causality of the associations between DNA methylation markers and cardiovascular outcomes, such as the recently published meta-analysis by Agha *et al*, genotyped the samples used in the EWASs to identify meQTLs [351]. The ARIES database only contains data for the intergenic cg21566642 and the *F2RL3*-regulator CpG we found. For both CpGs, only one genetic variant was available, so the genetic information available for the two-sample MR studies consisted of only one SNP. Our findings did not suggest a causal relationship between any of the two CpGs and CHD. However, cg21566642 was found to be differentially methylated in association with incident CVD, which suggest that demethylation at this CpG occurs before the event. Both CpGs were related to CVRFs, which suggests that their methylation status could mediate between CVR and cardiovascular outcomes.

As already mentioned in section 4.3.c, the CpG we found in *CPT1A*-regulator to be related to both prevalent CHD and incident CVD was previously shown to be associated with TG levels in both directions of the causal path, i.e. its methylation level may influence TG levels or be determined by them. This supports the idea that differential methylation at this CpG is determined prior to the cardiovascular outcome. Also, the prospective associations support the inference of causality for the CpGs in *AHRR*, *ABCG1*, *GARS*, *PARP9/DTX3L* and *H6PD*. All these non-conclusive findings underline the complex network connecting DNA methylation and other mechanisms regulating gene expression, CVRFs and cardiovascular outcomes.

Highlights of Objective 3

- Blood DNA methylation levels were associated with prevalent CHD and incident CVD using EWAS and candidate gene strategies
- Seventeen differentially methylated sites were found in total, many related to individual classical CVRFs
- Their genomic location supports their role as regulators of the expression of genes related to important processes in CVD, such as inflammation, metabolism, and multifunctional pathways
- Risk scores including these CpGs were associated with incident CVD independently of classical CVRFs. However, the Framingham risk function did not show any improvement in predictive capacity when these risk factors were included
- MR studies to assess the causality of the associations between the identified CpGs and clinical outcomes were not conclusive

4.5. Strengths and limitations

This thesis has several strengths and limitations that are common to the different objectives, as described below.

a) Strengths

1. **Large population-based studies.** The samples analyzed represent the general adult population, both men and women in the samples used for all three objectives of the thesis (REGICOR and FOS). All samples are large, ranging from hundreds to thousands of individuals, which increases the statistical power of our analyses and the validity of the results.

2. **Standardized and validated instruments, methods and procedures** have been used to collect and determine clinical, biological, sociodemographic, lifestyle and anthropometric variables.

3. **Bisulfite conversion-based methylation arrays and standardized quality controls, processing and analyses for DNA methylation studies.** We used a standardized analysis and quality control pipeline to analyze the association between DNA methylation and the traits of interest; to minimize false positive findings, the analyses included at least two populations.

4. **Blood samples as the DNA source.** Whole blood is one of the most easily accessible sample types, it requires less invasive methods, and has been proposed as a good proxy for methylation levels in other tissues [317,334,339]. Moreover, differences in leukocyte composition, which could confound changes in DNA methylation, have been shown to have only a very limited impact on EWAS findings [371].

5. **Established protocols for removing confounding.** In all analyses, we considered the effect of cell type heterogeneity [243] and unknown omics-related sources of confounding (surrogate variables) [245]. In some analyses we also adjusted for other confounder variables.

6. **Multiple approaches to address the objectives of this thesis.** These include different types of PA analyzed, an age-independent definition of CVR, complementary epigenome-wide and candidate gene association studies, and complementary approaches focused on prevalent and incident cardiovascular outcomes.

7. **Meta-analyses of the observed results.** We meta-analysed the results from the association studies, and assessed the consistency of the effect sizes and the direction of the associations across studies.

8. **Corrections for multiple comparisons and genomic inflation.** We applied a very conservative approach, namely the Bonferroni correction, to minimize false positive results. We also applied novel methodologies to reduce genomic inflation in EWAS analyses (bacon R function).

9. **Assessment of the clinical relevance of our results.** Some of the epigenetic biomarkers we identified were combined in risk scores to evaluate their potential as predictors of cardiovascular events.

b) Limitations

• Limitations in the study design and samples

1. **Blood samples as the DNA source.** Although the nature of blood makes it the preferred source for DNA methylation studies, some DNA methylation signatures are tissue- or cell-specific. We could miss those present in other important tissues related to the phenotypes of interest, such as muscle, adipose, or cardiac tissue.

2. **DNA methylation assessed at one time point.** Sample collection at multiple time points would enable us to elucidate DNA methylation changes over time.

3. **Lack of quantitative measurements of some of the exposures and covariates.** PA and smoking were self-reported through validated questionnaires. Biochemical traits such as lipid and glucose levels and BP were also self-reported

for the prevalent MI cases in objective 3 due to the acute nature of the event.

4. **Heterogeneity between the results observed in the different populations in each study**. We conducted fixed-effects meta-analyses, which does not allow for heterogeneity between studies. However, we assessed the consistency of the effect sizes and the direction of the associations.

5. **Populations from Western countries.** Individuals with different ethnic origins may show differences in the DNA methylation markers we found.

6. **Difficulty in determining causality.** The case-control and cross-sectional designs preclude causal inference. Therefore, we used MR studies to assess causality in objective 3, although this study design also has some limitations (see below).

7. **Lack of AMI information in other populations.** In the Framingham population, information about the incidence of CHD events was not further stratified into subtypes of events such as AMI, unstable angina, and stable angina. Therefore, we did not have access to information about AMI events in all populations, in order to compare methylation changes related to this more severe outcome.

- **Analytical and technical limitations**

1. **Use of 450k versus EPIC arrays.** One third of the CpGs analysed in the REGICOR case-control could not be assessed in the FOS population because the methylation arrays used had different numbers of interrogated CpGs.

2. **DNA methylation measured using β-values instead of M-values in objective 3.** M-values better identify differentially methylated CpG sites than β-values, which are the measures currently recommended by Illumina. However, the analyses using M- or β-values provided similar and consistent results.

3. **Inflation in the results of the WHI sample with CHD data from objective 3.** We could not reduce this inflation using the

bacon R package. However, the CpGs identified in this sample were also related to CHD in the FOS population.

4. **Unavailability of genotype data for the populations included in this thesis.** We intended to perform MR studies to address causality, but there is still limited availability of meQTL datasets. The lack of valid genetic instrumental variables has hampered our ability to assess the causality of the associations discovered.

5. **Batch effect.** We used appropriate methods to correct for non-biological sources of variation, such as methylation data standardization by batch, removal of outlier values, using robust multivariate models, or adjusting for surrogate variables. However, technical heterogeneity is an intrinsic limitation of studies with high-throughput data. While several methods have been proposed to remove or minimize this heterogeneity, there is no consensus on which is the most accurate, and it may vary depending on the study. Also, biological heterogeneity can be confused with batch effects and removed in error [372].

5. FUTURE PERSPECTIVES

Illustration by Evelien Jagtman

"The future belongs to those who believe in the beauty of their dreams."

Eleanor Roosevelt (1884-1962).

To better understand the interplay between DNA methylation, cardiovascular risk and cardiovascular outcomes, the following questions should be prioritized in further research:

1. ***What about potential DNA methylation patterns that cannot be tested in a second sample?***

   450k arrays are no longer commercialized, but much of the available DNA methylation data has been generated using them, which prevents researchers from validating data obtained from EPIC arrays (explained in section 1.4.d). This limitation may be solved by methods that impute DNA methylation levels at specific sites based on neighbouring data (similar to those used in GWAS analyses), or that at least speed up the validation of potential loci.

2. ***Is differential DNA methylation a major driver of cardiovascular and coronary outcomes? Or is it a "passenger" instead of a driver of the disease?***

   To date there have been few MR studies of the causal association between DNA methylation loci and different traits, generally because of the scarcity of samples with both DNA methylation data and genotype data. In this regard, it is essential to identify meQTLs and then share these data via a database that is openly accessible to the scientific community.

3. ***Are PA-related loci also associated with incident cardiovascular events?***

   We have only addressed the association between PA and DNA methylation, but have not assessed the clinical value that the identified signatures may provide to support the benefits of PA for health. It is necessary to assess their potential as predictors of incident CVD.

4. ***Does differential DNA methylation play an intermediate role between lifestyle or environmental factors and cardiovascular outcomes?***

MR studies would provide insights into the causality of CpG-CVD associations. Also, analyses of the association between CpG sites and individual CVRFs inform us about the role of DNA methylation as a potential intermediate mechanism underlying the effects of CVRFs on cardiovascular health. This scenario is plausible for exposure variables such as smoking, PA or environmental pollutants. However, it could be that DNA methylation causally determines a specific CVRF and that this influences cardiovascular health through a different molecular mechanism. Integrating causation and mediation analyses would provide more insights into this complex interplay.

5. ***What is the relationship between DNA methylation and other lifestyle, environmental, and physiological factors?***

Little is known about the impact of diet on DNA methylation. Also, there is growing research into patterns of DNA methylation signatures that arise following prenatal exposure to various factors. All of these potential associations could also be important for cardiovascular outcomes. Thus, there is a need for further epidemiological and epigenomic research on these factors.

6. ***What is the relationship between DNA methylation and other cardio-metabolic traits?***

The associations between DNA methylation and conditions such as diabetes, metabolic syndrome, hypertension, and subclinical atherosclerosis, are not well understood yet. In the REGICOR study, we are currently addressing some of these associations.

7. ***How can we estimate cardiovascular risk more accurately?***

There have been few, and unsuccessful, efforts to assess the utility of new biomarkers in clinical settings by including them in CVR functions. Different approaches to estimating CVR should be explored. Machine learning algorithms that combine data from different layers of information and different populations and

samples have promising potential for individualized preventive medicine.

8. *Are DNA methylation-based predictors of age useful for predicting cardiovascular outcomes?*

Several DNA-methylation-based predictors of age have been developed, such as the epigenetic age estimators proposed by Horvath [373] and Hannum *et al* [374]. These authors have studied the associations between these predictors and CVR and cardiovascular outcomes [250,375,376]. In the REGICOR study, we have conducted a preliminary analysis in which we included these predictors in CVR functions, with negative results. However, further efforts based on other estimators are required.

9. *Which DNA methylation signatures are associated with the assessed phenotypes in subgroups of the general population?*

All of our studies have been conducted in the general adult population. However, older individuals may have different DNA methylation loci associated with different exposures, CVRFs, CVR, and cardiovascular outcomes. Moreover, population subgroups with cardio-metabolic diseases such as diabetes, hypertension, dyslipidaemia, and obesity may experience different methylome changes. Finally, in individuals with other diseases such as cancer, mental disorders, or immunodeficiency, the DNA methylation loci that are associated with cardiovascular health may be different. Further research in different subgroups is needed.

10. *Which specific DNA methylation patterns are disease drivers and where are they located in the genome?*

The CpG sites identified should be interpreted at the genomic information level in order to better understand DNA methylation mechanisms, i.e. whether the causal CpG sites are located within CGIs, whether these map to a promoter region, etc. While differentially methylated CpG sites involved in the causal pathway to cardiovascular outcomes may lie anywhere in the genome, they

probably tend to map to certain genomic contexts. In this regard, analysing differentially methylated regions may complement the information provided by single-nucleotide sites. This would require implementation and standardization of the methods.

11. *Which patterns of differential DNA methylation are inherited in cardiovascular outcomes?*

The methylation or demethylation of a specific loci may not only be triggered by modifiable factors, but may also be inherited and transmitted to the next generation. There is a need for new methods to distinguish inherited methylome marks from those that are dependent on the environment.

12. *Do those differential patterns correlate with biological changes in other layers of molecular information?*

Deciphering DNA methylation patterns associated with cardiovascular health is only one layer of the complex underlying molecular network. Integrating these patterns with those identified in other layers of biological information would provide a broader picture of what is taking place within our cells in the non-disease state, and what changes in the disease state. Apart from DNA methylation, other important layers include genetics, chromatin/histone changes, ncRNA profiles, transcriptome, proteome, and metabolome. Integrating these layers requires new and standardized methodology.

13. *Can DNA methylation loci be influenced by the microbiome and the virome? What about acute infections?*

To understand the biological networks underlying cardiovascular diseases, in addition to integrating molecular mechanisms and traits, we also need data on the microbiome and virome in order to clarify this complex scenario. In addition, acute infections may affect epigenetic mechanisms that modulate CVR, especially infection by microorganisms that require components of the host cell (such as RNA viruses).

14. *Are DNA methylation loci potential therapeutic targets?*

As for the previous question, the most interesting loci that could be targeted for therapeutic purposes would be those with biological outcomes that are consistent between different information layers. At this point, MR studies are required to disentangle the direction of the associations, in order to consider or discard a biomarker as a potential therapeutic target.

6. CONCLUSIONS

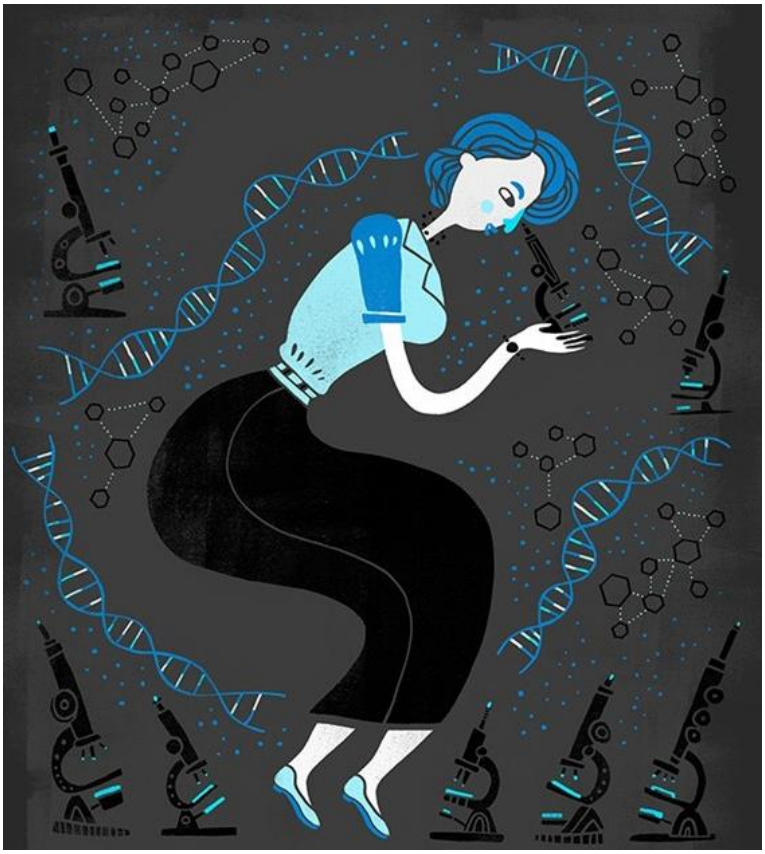


Illustration by Rachel Ignotofsky, modified

(Source: Women in Science: Fifty Fearless Pioneers Who Changed the World)

"If we assume we have arrived, we stop searching, we stop developing."

Jocelyn Bell Burnell (1943).

a) General conclusion 1

We identified differential DNA methylated signatures in blood cells associated with a lifestyle factor: physical activity.

- **Specific conclusions**

a.1) We report two CpGs, cg24155427 and cg09565397, that showed differential methylation levels in association with leisure-time physical activity.

a.2) These two CpGs showed a non-linear dose-response relationship with moderate- and vigorous-intensity physical activity, which is recommended by current health guidelines. Specifically, at high levels of physical activity, higher doses of moderate-vigorous physical activity were associated with lower methylation levels at those CpGs.

a.3) cg09565397 maps to *DGAT1*, which is involved in triglyceride metabolism. Therefore, this finding supports the impact of physical activity on lipid homeostasis, with DNA methylation as a potential intermediate mechanism.

b) General conclusion 2

We identified differential DNA methylation patterns in blood cells related to the cardiovascular risk factors load estimated by cardiovascular risk independent of age.

- **Specific conclusions**

b.1) We report eight CpGs that showed hypomethylation related to age-independent cardiovascular risk.

b.2) All eight CpGs were also related to classical cardiovascular risk factors. All but one were related to smoking, supporting its impact on both DNA methylation and cardiovascular risk. Four were associated with body mass index, two as novel findings.

b.3) These CpGs map to three intergenic regions, and to the genes *AHRR*, *CPT1A*, *PPIF*, and *SBNO2*, which are involved in processes such as inflammation, lipid metabolism, and multifunctional pathways.

b.4) Two CpGs were located in the same genetic region, so we combined seven of them into CpG-based risk scores, which were predictive of incident CVD independently of classical cardiovascular risk factors. The risk scores were also associated with measures of subclinical atherosclerosis.

c) General conclusion 3

We identified differential DNA methylation patterns in blood cells associated with cardiovascular and coronary heart diseases and evaluated their validity as predictive biomarkers of incident CVD.

• Specific conclusions

c.1) Via a systematic review of this topic, we identified loci with differential DNA methylation related to cardiovascular risk, namely 52 genes with methylation levels with the same reported direction of association in at least two epigenome-wide association studies. These genes can now be included in candidate-gene strategies to analyse their association with cardiovascular outcomes.

c.2) We report seventeen CpGs that had differential methylation levels related to prevalent coronary heart disease and incident cardiovascular and coronary heart disease. Some of these CpGs are novel findings and map to *C1QTNF7*, *RFTN1*, *ADPGK*, *DSCAML1*, *GARS* and *H6PD*.

c.3) Most of these CpGs were also associated with classical cardiovascular risk factors. Some others highlight processes such as inflammation and multifunctional pathways.

c.4) Based on the results of our analysis, we constructed CpG-based risk scores, which we found to be related to incident CVD independent of classical cardiovascular risk factors. However, these scores did not improvement the capacity of risk functions to predictive cardiovascular risk.

c.5) The results of the Mendelian Randomisation studies performed to assess the causality of the associations were non-conclusive.

7. BIBLIOGRAPHY

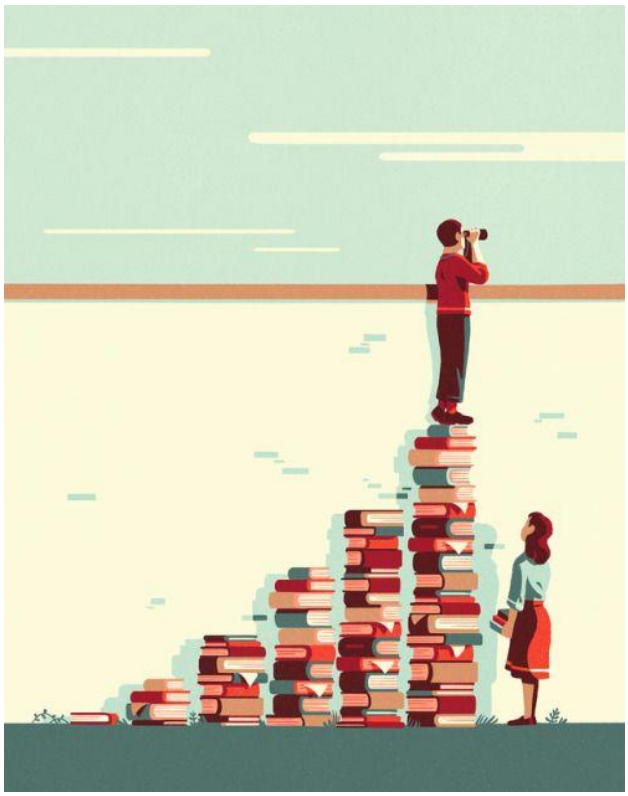


Illustration by Tom Haugomat

"For most of history, anonymous was a woman."

Virginia Woolf (1882–1941).

1. Roth GA, Abate D, Abate KH, *et al*. Global, regional, and national age-sex-specific mortality for 282 causes of death in 195 countries and territories, 1980-2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet. 2018;392(10159):1736-88.

2. James SL, Abate D, Abate KH, *et al*. Global, Regional, and National Incidence, Prevalence, and Years Lived with Disability for 354 Diseases and Injuries for 195 Countries and Territories, 1990-2017: A Systematic Analysis for the Global Burden of Disease Study 2017. Lancet. 2018;392 (10159):1789-858.

3. Kyu HH, Abate D, Abate KH, *et al*. Global, Regional, and National Disability-Adjusted Life-Years (DALYs) for 359 Diseases and Injuries and Healthy Life Expectancy (HALE) for 195 Countries and Territories, 1990-2017: A Systematic Analysis for the Global Burden of Disease Study 2017. 2018;392(10159):1859-922.

4. Mahmood SS, Levy D, Vasan RS, Wang TJ. The Framingham Heart Study and the epidemiology of cardiovascular disease: a historical perspective. Lancet. 2014;383(9921):999-1008.

5. Dawber TR, Meadors GF, Moore FE. Epidemiological approaches to heart disease: the Framingham Study. Am J Public Health Nations Health. 1951;41(3):279-81.

6. Kannel WB, Dawber TR, Kagan A, *et al*. Factors of risk in the development of coronary heart disease - six year follow-up experience. The Framingham Study. Ann Intern Med. 1961;55(1):33-50.

7. Truett J, Cornfield J, Kannel W. A multivariate analysis of the risk of coronary heart disease in Framingham. J Chronic Dis. 1967;20(7):511-24.

8. Stanaway JD, Afshin A, Gakidou E, *et al*. Global, regional, and national comparative risk assessment of 84 behavioural, environmental and occupational, and metabolic risks or clusters of risks for 195 countries and territories, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet. 2018;392(10159):1923-94.

9. United Nations Department of Economic and, Social Affairs: Population Division. World Population Prospects. The 2017 Revision. Retrieved from: https://www.un.org/development/desa/publications/world-population-prospects-the-2017-revision.html.

10. Degano IR, Elosua R, Marrugat J. Epidemiology of Acute Coronary Syndromes in Spain: Estimation of the Number of Cases and Trends From 2005 to 2049. Rev Española Cardiol. 2013;66(6):472-81.

11. Centers for Disease Control and Prevention. Achievements in Public Health, 1900-1999: Changes in the Public Health System. JAMA. 2000;283(6):735-8.

12. Honda S, Asaumi Y, Yamane T, *et al*. Trends in the clinical and pathological characteristics of cardiac rupture in patients with acute myocardial infarction over 35 years. J Am Heart Assoc. 2014;3(5):e000984.

13. McSweeney JC, Cody M, O'Sullivan P, *et al*. Women's early warning symptoms of acute myocardial infarction. Circulation. 2003;108(21):2619-23.

14. de Torbal A, Boersma E, Kors JA, *et al*. Incidence of recognized and unrecognized myocardial infarction in men and women aged 55 and older: the Rotterdam Study. Eur Heart J. 2006;27(6):729-36.

15. Thygesen K, Alpert JS, Jaffe AS, *et al*. Fourth universal definition of myocardial infarction (2018). Eur Heart J. 2019;40(3):237-69.

16. Sandoval Y, Apple FS, Smith SW. High-sensitivity cardiac troponin assays and unstable angina. Eur Heart J Acute Cardiovasc Care. 2018;7(2):120-28.

17. Rousan TA, Thadani U. Stable Angina Medical Therapy Management Guidelines: A Critical Review of Guidelines from the European Society of Cardiology and National Institute for Health and Care Excellence. Eur Cardiol Rev. 2019;14(1):18-22.

18. Ross R. Atherosclerosis - An Inflammatory Disease. N Engl J Med. 1999;340(2):115-26.

19. Napoli C, D'Armiento FP, Mancini FP, *et al.* Fatty streak formation occurs in human fetal aortas and is greatly enhanced by maternal hypercholesterolemia. Intimal accumulation of low density lipoprotein and its oxidation precede monocyte recruitment into early atherosclerotic lesions. J Clin Invest. 1997;100(11):2680-90.

20. Sanz J, Fayad ZA. Imaging of atherosclerotic cardiovascular disease. Nature. 2008;451(7181):953-57.

21. Stary HC, Chandler AB, Dinsmore RE, *et al.* A definition of advanced types of atherosclerotic lesions and a histological classification of atherosclerosis. A report from the Committee on Vascular Lesions of the Council on Arteriosclerosis, American Heart Association. Circulation. 1995;92(5):1355-74.

22. Wingard DL, Suarez L, Barrett-Connor E. The Sex Differential In Mortality From All Causes And Ischemic Heart Disease. Am J Epidemiol. 1983;117(2):165-72.

23. Dawber TR, Kannel WB. The Framingham study. An epidemiological approach to coronary heart disease. Circulation. 1966;34(4):553-5.

24. Ford ES, Ajani UA, Croft JB, *et al.* Explaining the Decrease in U.S. Deaths from Coronary Disease, 1980–2000. N Engl J Med. 2007;356(23):2388-98.

25. Roberto CA, Swinburn B, Hawkes C, *et al.* Patchy progress on obesity prevention: emerging examples, entrenched barriers, and new thinking. Lancet. 2015;385(9985):2400-09.

26. Sniderman AD, Furberg CD. Age as a modifiable risk factor for cardiovascular disease. Lancet. 2008;371(9623):1547-49.

27. Marrugat J, Subirana I, Ramos R, *et al.* Derivation and validation of a set of 10-year cardiovascular risk predictive functions in Spain: The FRESCO Study. Prev Med. 2014;61:66-74.

28. Sans S, Fitzgerald AP, Royo D, Conroy R, Graham I. Calibrating the SCORE Cardiovascular Risk Chart for Use in Spain. Rev Española Cardiol. 2007;60(5):476-85.

29. Arnett DK, Blumenthal RS, Albert MA, *et al.* 2019 ACC/AHA Guideline on the Primary Prevention of Cardiovascular Disease: A Report of the American College of

Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. Circulation. 2019;140(11):e596-e646.

30. Piepoli MF, Hoes AW, Agewall S, *et al.* 2016 European Guidelines on cardiovascular disease prevention in clinical practice. Eur Heart J. 2016;37(29):2315-81.

31. Singh RB, Mengi SA, Xu Y-J, *et al.* Pathogenesis of atherosclerosis: A multifactorial process. Exp Clin Cardiol. 2002;7(1):40-53.

32. North BJ, Sinclair DA. The intersection between aging and cardiovascular disease. Circ Res. 2012;110(8):1097-108.

33. D'Agostino RB, Vasan RS, Pencina MJ, *et al.* General Cardiovascular Risk Profile for Use in Primary Care: the Framingham Heart Study. Circulation. 2008;117(6):743-53.

34. Elosua R. Cardiovascular risk functions: usefulness and limitations. Rev Esp Cardiol. 2014;67(2):77-9.

35. Groenewegen K, den Ruijter H, Pasterkamp G, *et al.* Vascular age to determine cardiovascular disease risk: A systematic review of its concepts, definitions, and clinical applications. Eur J Prev Cardiol. 2016;23(3):264-74.

36. Appelman Y, van Rijn BB, Ten Haaf ME, *et al.* Sex differences in cardiovascular risk factors and disease prevention. Atherosclerosis. 2015;241(1):211-18.

37. Gillum RF, Mehari A, Curry B, Obisesan TO. Racial and geographic variation in coronary heart disease mortality trends. BMC Public Health. 2012;12(1):410.

38. Muntner P, Abdalla M, Correa A, *et al.* Hypertension in Blacks. Hypertension. 2017;69(5):761-9.

39. Mensah GA, Mokdad AH, Ford ES, *et al.* State of Disparities in Cardiovascular Health in the United States. Circulation. 2005;111(10):1233-41.

40. Cainzos-Achirica M, Fedeli U, Sattar N, *et al.* Epidemiology, risk factors, and opportunities for prevention of cardiovascular disease in individuals of South Asian ethnicity living in Europe. Atherosclerosis. 2019;286:105-13.

41. Jonas MA, Oates JA, Ockene JK, Hennekens CH. Statement on smoking and cardiovascular disease for health care professionals. American Heart Association. Circulation. 1992;86(5):1664-9.

42. Law MR, Wald NJ. Environmental tobacco smoke and ischemic heart disease. Prog Cardiovasc Dis. 2003;46(1):31-8.

43. Agüero F, Dégano IR, Subirana I, *et al*. Impact of a partial smoke-free legislation on myocardial infarction incidence, mortality and case-fatality in a population-based registry: the REGICOR Study. PLoS One. 2013;8(1):e53722.

44. Gordon T, Kannel WB, McGee D, Dawber TR. Death and coronary attacks in men after giving up cigarette smoking. A report from the Framingham study. Lancet. 1974;2(7893):1345-8.

45. Kuller LH, Ockene JK, Meilahn E, *et al*. Cigarette smoking and mortality. MRFIT Research Group. Prev Med. 1991;20(5):638-54.

46. Lightwood JM, Glantz SA. Short-term economic and health benefits of smoking cessation: myocardial infarction and stroke. Circulation. 1997;96(4):1089-96.

47. Fa-Binefa M, Clará A, Pérez-Fernández S, *et al*. Early smoking-onset age and risk of cardiovascular disease and mortality. Prev Med. 2019;124:17-22.

48. Barnoya J, Glantz SA. Cardiovascular effects of secondhand smoke: nearly as large as smoking. Circulation. 2005;111(20):2684-98.

49. Mamun AA, O'Callaghan MJ, Williams GM, Najman JM. Maternal Smoking During Pregnancy Predicts Adult Offspring Cardiovascular Risk Factors - Evidence from a Community-Based Large Birth Cohort Study. Li S, ed. PLoS One. 2012;7(7):e41106.

50. Huxley RR, Woodward M. Cigarette smoking as a risk factor for coronary heart disease in women compared with men: a systematic review and meta-analysis of prospective cohort studies. Lancet. 2011;378(9799):1297-305.

51. Lidón-Moyano C, Fu M, Pérez-Ortuño R, *et al*. Impact of the Spanish smoking legislation among smokers: A longitudinal

study with biomarkers in Barcelona (Spain). Addict Behav. 2018;87:101-108.

52. Fernández E, Fu M, Pérez-Ríos M, *et al.* Changes in Secondhand Smoke Exposure After Smoke-Free Legislation (Spain, 2006–2011). Nicotine Tob Res. 2017;19(11):1390-94.

53. Berglund L, Brunzell JD, Goldberg AC, *et al.* Evaluation and treatment of hypertriglyceridemia: An endocrine society clinical practice guideline. J Clin Endocrinol Metab. 2012;97(9):2969-89.

54. Saeed A, Feofanova E V., Yu B, *et al.* Remnant-Like Particle Cholesterol, Low-Density Lipoprotein Triglycerides, and Incident Cardiovascular Disease. J Am Coll Cardiol. 2018;72(2):156-69.

55. Reiner Ž. Hypertriglyceridaemia and risk of coronary artery disease. Nat Rev Cardiol. 2017;14(7):401-411.

56. Dron JS, Hegele RA. Genetics of Triglycerides and the Risk of Atherosclerosis. Curr Atheroscler Rep. 2017;19(7):31.

57. Holmes M.V., Asselbergs FW, Palmer TM, *et al.* Mendelian randomization of blood lipids for coronary heart disease. Eur Heart J. 2015;36(9):539-50.

58. Klag MJ, Ford DE, Mead LA, *et al.* Serum Cholesterol in Young Men and Subsequent Cardiovascular Disease. N Engl J Med. 1993;328(5):313-18.

59. Grundy SM, Stone NJ, Bailey AL, *et al.* 2018 AHA/ACC/AACVPR/AAPA/ABC/ACPM/ADA/AGS/APhA/ASPC/NLA/PCNA Guideline on the Management of Blood Cholesterol: A Report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines.. J Am Coll Cardiol. 2019;73(24):e285-e350.

60. Gencer B, Mach F. Lipid management in ACS: Should we go lower faster? Atherosclerosis. 2018;275:368-75.

61. Wang HH, Garruti G, Liu M, *et al.* Cholesterol and Lipoprotein Metabolism and Atherosclerosis: Recent Advances In reverse Cholesterol Transport. Ann Hepatol. 2017;16(1):s27-s42.

62. Rader DJ, Hovingh GK. HDL and cardiovascular disease. Lancet. 2014;384(9943):618-25.

63. Rohatgi A, Khera A, Berry JD, *et al*. HDL Cholesterol Efflux Capacity and Incident Cardiovascular Events. N Engl J Med. 2014;371(25):2383-93.

64. Keene D, Price C, Shun-Shin MJ, Francis DP. Effect on cardiovascular risk of high density lipoprotein targeted drug treatments niacin, fibrates, and CETP inhibitors: meta-analysis of randomised controlled trials including 117,411 patients. BMJ. 2014;349:g4379.

65. Voight BF, Peloso GM, Orho-Melander M, *et al*. Plasma HDL cholesterol and risk of myocardial infarction: a mendelian randomisation study. Lancet. 2012;380(9841):572-80.

66. Hernáez Á, Castañer O, Elosua R, *et al*. Mediterranean Diet Improves High-Density Lipoprotein Function in High-Cardiovascular-Risk Individuals. Circulation. 2017;135(7):633-43.

67. O'Donnell CJ, Elosua R. Cardiovascular risk factors. Insights from Framingham Heart Study. Rev Esp Cardiol. 2008;61(3):299-310.

68. Goff DC, Lloyd-Jones DM, Bennett G, *et al*. 2013 ACC/AHA Guideline on the Assessment of Cardiovascular Risk: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. Circulation. 2014;129(25):S74-5.

69. Gillis EE, Sullivan JC. Sex Differences in Hypertension: Recent Advances. Hypertension. 2016;68(6):1322-27.

70. Reuter H, Jordan J. Status of hypertension in Europe. Curr Opin Cardiol. 2019;34(4):342-49.

71. Lee SI, Patel M, Jones CM, Narendran P. Cardiovascular disease and type 1 diabetes: Prevalence, prediction and management in an ageing population. Ther Adv Chronic Dis. 2015;6(6):347-74.

72. Strain WD, Paldánius PM. Diabetes, cardiovascular disease and the microcirculation. Cardiovasc Diabetol. 2018;17(1):57.

73. Peters SAE, Huxley RR, Woodward M. Diabetes as a risk factor for stroke in women compared with men: a systematic

review and meta-analysis of 64 cohorts, including 775,385 individuals and 12,539 strokes. Lancet. 2014;383(9933):1973-80.

74. Tuomilheto J, Lindström J, Eriksson JG, *et al.* Prevention of Type 2 Diabetes Mellitus By Changes in Lifestyle Among Subjects With Impaired Glucose Tolerance. N Engl J Med. 2001;344(18):1343-50.

75. Diabetes Prevention Program Research G, Knowler WC, Fowler SE, *et al.* 10-year follow-up of diabetes incidence and weight loss in the Diabetes Prevention Program Outcomes Study. Lancet. 2009;374(9702):1677-86.

76. Knowler WC, Barrett-Connor E, Fowler SE, *et al.* Reduction In The Incidence Of Type 2 Diabetes With Lifestyle Intervention Or Metformin. N Engl J Med. 2002;346(6):393-403.

77. Parto P, Lavie CJ. Obesity and Cardiovascular Diseases. Curr Probl Cardiol. 2017;42(11):376-94.

78. Després JP, Lemieux I, Bergeron J, *et al.* Abdominal obesity and the metabolic syndrome: contribution to global cardiometabolic risk. Arterioscler Thromb Vasc Biol. 2008;28(6):1039-49.

79. Elagizi A, Kachur S, Lavie CJ, *et al.* An Overview and Update on Obesity and the Obesity Paradox in Cardiovascular Diseases. Prog Cardiovasc Dis. 2018;61(2):142-150.

80. O'Neill S, O'Driscoll L. Metabolic syndrome: a closer look at the growing epidemic and its associated pathologies. Obes Rev. 2015;16(1):1-12.

81. Alberti KG, 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; National Heart, Lung, and Blood Institute; American Heart Association; World Heart Federation; International Atherosclerosis Society; and International Association for the Study of Obesity. Circulation. 2009;120(16):1640-5.

82. Mottillo S, Filion KB, Genest J, *et al.* The metabolic syndrome and cardiovascular risk: A systematic review and meta-analysis. J Am Coll Cardiol. 2010;56(14):1113-32.

83. Elosua R, Lluís-Ganella C, Subirana I, *et al.* Cardiovascular risk factors and ischemic heart disease: Is the Confluence of Risk Factors Greater Than the Parts? A Genetic Approach. Circ Cardiovasc Genet. 2016;9(3):279-86.

84. Hernáez Á, Zomeño MD, Dégano IR, *et al.* Excess Weight in Spain: Current Situation, Projections for 2030, and Estimated Direct Extra Cost for the Spanish Health System. Rev Esp Cardiol. 2018. pii: S1885-5857(18)30440-7.

85. Neufer PD, Bamman MM, Muoio DM, *et al.* Understanding the Cellular and Molecular Mechanisms of Physical Activity-Induced Health Benefits. Cell Metab. 2015;22(1):4-11.

86. Fiuza-Luces C, Santos-Lozano A, Joyner M, *et al.* Exercise benefits in cardiovascular disease: beyond attenuation of traditional risk factors. Nat Rev Cardiol. 2018;15(12):731-43.

87. Kahlmeier S, Wijnhoven TMA, Alpiger P, *et al.* National physical activity recommendations: systematic overview and analysis of the situation in European countries. BMC Public Health. 2015;15(1):133.

88. Kannel WB. Habitual level of physical activity and risk of coronary heart disease: the Framingham study. Can Med Assoc J. 1967;96(12):811-2.

89. Paffenbarger RS, Hale WE. Work activity and coronary heart mortality. N Engl J Med. 1975;292(11):545-50.

90. Wahid A, Manek N, Nichols M, *et al.* Quantifying the Association Between Physical Activity and Cardiovascular Disease and Diabetes: A Systematic Review and Meta-Analysis. J Am Heart Assoc. 2016;5(9):e002495.

91. Sattelmair J, Pertman J, Ding EL, *et al.* Dose Response Between Physical Activity and Risk of Coronary Heart Disease: a Metaanalysis. Circulation. 2011;124(7):789-95.

92. Backshall J, Ford GA, Bawamia B, *et al.* Physical activity in the management of patients with coronary artery disease: a review. Cardiol Rev. 23(1):18-25.

93. Gomes-Neto M, Durães AR, Reis HFCD, *et al.* High-intensity interval training versus moderate-intensity continuous training on exercise capacity and quality of life in patients with coronary artery disease: A systematic review and meta-analysis. Eur J Prev Cardiol. 2017;24(16):1696-707.

94. Patel A V., Bernstein L, Deka A, *et al.* Leisure time spent sitting in relation to total mortality in a prospective cohort of US adults. Am J Epidemiol. 2010;172(4):419-29.

95. Warren TY, Barry V, Hooker SP, *et al.* Sedentary behaviors increase risk of cardiovascular disease mortality in men. Med Sci Sports Exerc. 2010;42(5):879-85.

96. Katzmarzyk PT, Church TS, Craig CL, Bouchard C. Sitting time and mortality from all causes, cardiovascular disease, and cancer. Med Sci Sports Exerc. 2009;41(5):998-1005.

97. Matthews CE, George SM, Moore SC, *et al.* Amount of time spent in sedentary behaviors and cause-specific mortality in US adults. Am J Clin Nutr. 2012;95(2):437-45.

98. Chomistek AK, Manson JE, Stefanick ML, *et al.* Relationship of sedentary behavior and physical activity to incident cardiovascular disease: results from the Women's Health Initiative. J Am Coll Cardiol. 2013;61(23):2346-54.

99. Orkaby AR, Forman DE. Physical activity and CVD in the elderly: an expert's perspective. Expert Rev Cardiovasc Ther. 2018 Jan;16(1):1-10.

100. Ekelund U, Luan J, Sherar LB, *et al.* Moderate to vigorous physical activity and sedentary time and cardiometabolic risk factors in children and adolescents. JAMA. 2012;307(7):704-12.

101. Cárdenas Fuentes G, Bawaked RA, Martínez González MÁ, *et al.* Association of physical activity with body mass index, waist circumference and incidence of obesity in older adults. Eur J Public Health. 2018;28(5):944-50.

102. Curtis GL, Chughtai M, Khlopas A, *et al.* Impact of Physical Activity in Cardiovascular and Musculoskeletal Health: Can Motion Be Medicine? J Clin Med Res. 2017;9(5):375-81.

103. Thompson PD, Franklin BA, Balady GJ, *et al.* Exercise and acute cardiovascular events placing the risks into perspective: a

scientific statement from the American Heart Association Council on Nutrition, Physical Activity, and Metabolism and the Council on Clinical Cardiology. Circulation. 2007;115(17):2358-68.

104. Dinu M, Pagliai G, Casini A, Sofi F. Mediterranean diet and multiple health outcomes: an umbrella review of meta-analyses of observational studies and randomised trials. Eur J Clin Nutr. 2018;72(1):30-43.

105. Estruch R, Ros E, Salas-Salvadó J, *et al.* Primary Prevention of Cardiovascular Disease with a Mediterranean Diet Supplemented with Extra-Virgin Olive Oil or Nuts. N Engl J Med. 2018;378(25):e34.

106. Tang WHW, Kitai T, Hazen SL. Gut Microbiota in Cardiovascular Health and Disease. Circ Res. 2017;120(7):1183-96.

107. Mostofsky E, Chahal HS, Mukamal KJ, *et al.* Alcohol and Immediate Risk of Cardiovascular Events: A Systematic Review and Dose-Response Meta-Analysis. Circulation. 2016;133(10):979-87.

108. Schultz WM, Kelli HM, Lisko JC, *et al.* Socioeconomic Status and Cardiovascular Outcomes. Circulation. 2018;137(20):2166-78.

109. Ginty AT, Kraynak TE, Fisher JP, Gianaros PJ. Cardiovascular and autonomic reactivity to psychological stress: Neurophysiological substrates and links to cardiovascular disease. Auton Neurosci. 2017;207:2-9.

110. Khan MS, Aouad R. The Effects of Insomnia and Sleep Loss on Cardiovascular Disease. Sleep Med Clin. 2017;12(2):167-77.

111. Fiordelisi A, Piscitelli P, Trimarco B, Coscioni E, Iaccarino G, Sorriento D. The mechanisms of air pollution and particulate matter in cardiovascular diseases. Heart Fail Rev. 2017;22(3):337-47.

112. Gansevoort RT, Correa-Rotter R, Hemmelgarn BR, *et al.* Chronic kidney disease and cardiovascular risk: epidemiology, mechanisms, and prevention. Lancet. 2013;382(9889):339-52.

113. Graham I, Atar D, Borch-Johnsen K, *et al.* European guidelines on cardiovascular disease prevention in clinical

practice: executive summary: Fourth Joint Task Force of the European Society of Cardio. Eur Heart J. 2007;14 (2):E1-40.

114. Wilson PW, D'Agostino RB, Levy D, *et al*. Prediction of coronary heart disease using risk factor categories. Circulation. 1998;97(18):1837-47.

115. Grau M, Marrugat J. Risk Functions and the Primary Prevention of Cardiovascular Disease. Rev Esp Cardiol. 2008;61(4):404-16.

116. Marrugat J, D'Agostino R, Sullivan L, *et al*. An adaptation of the Framingham coronary heart disease risk function to European Mediterranean areas. J Epidemiol Community Health. 2003;57(8):634-8.

117. D'Agostino, RB Sr, Grundy S, Sullivan LM, *et al*. Validation of the Framingham coronary heart disease prediction scores: Results of a Multiple Ethnic Groups Investigation.. JAMA. 2001;286(2):180-7.

118. Liu J, Hong Y, D'Agostino RB, *et al*. Predictive value for the Chinese population of the Framingham CHD risk assessment tool compared with the Chinese Multi-Provincial Cohort Study. JAMA. 2004;291(21):2591-99.

119. Brotons C, Lobos JM, Royo-Bordonada MÁ, *et al*. Implementation of Spanish adaptation of the European guidelines on cardiovascular disease prevention in primary care. BMC Fam Pract. 2013;14:36.

120. Royo-Bordonada MÁ, Lobos JM, Brotons C, *et al*. El estado de la prevención cardiovascular en España. [The condition of the cardiovascular prevention in Spain]. Med Clin. 2014;142(1):7-14.

121. Royo-Bordonada MÁ, Armario P, Lobos Bejarano JM, *et al*. Adaptación española de las guías europeas de 2016 sobre prevención de la enfermedad cardiovascular en la práctica clínica. Semergen. 2017;43(4):295-311.

122. Conroy RM, Pyörälä K, Fitzgerald AP, *et al*. Estimation of ten-year risk of fatal cardiovascular disease in Europe: The SCORE project. Eur Heart J. 2003;24(11):987-1003.

123. Baena-Díez JM, Subirana I, Ramos R, *et al*. Validity Assessment of Low-risk SCORE Function and SCORE

Function Calibrated to the Spanish Population in the FRESCO Cohorts. Rev Esp Cardiol. 2018;71(4):274-82.

124. Marrugat J, Subirana I, Comín E, *et al.* Validity of an adaptation of the Framingham cardiovascular risk function: the VERIFICA Study. J Epidemiol Community Health. 2007;61(1):40-7.

125. Marrugat J, Vila J, Baena-Díez JM, *et al.* Relative Validity of the 10-Year Cardiovascular Risk Estimate in a Population Cohort of the REGICOR Study. Rev Española Cardiol. 2011;64(5):385-94.

126. Baena-Díez JM, Grau M, Sánchez-Pérez R, *et al.* The REGICOR-calibrated function provides a better classification of high-risk patients on statin treatment in the Spanish population than the Framingham or SCORE classifications. Rev Esp Cardiol. 2009;62(10):1134-40.

127. Marrugat J, Solanas P, D'Agostino R, *et al.* Coronary risk estimation in Spain using a calibrated Framingham function. Rev Esp Cardiol. 2003;56(3):253-61.

128. Hlatky MA, Greenland P, Arnett DK, *et al.* Criteria for Evaluation of Novel Markers of Cardiovascular Risk. Circulation. 2009;119(17):2408-16.

129. Vlachopoulos C, Xaplanteris P, Aboyans V, *et al.* The role of vascular biomarkers for primary and secondary prevention. A position paper from the European Society of Cardiology Working Group on peripheral circulation: Endorsed by the Association for Research into Arterial Structure and Physiology (ARTERY) Society. Atherosclerosis. 2015;241(2):507-32.

130. Steyerberg EW, Vickers AJ, Cook NR, *et al.* Assessing the performance of prediction models: a framework for traditional and novel measures. Epidemiology. 2010;21(1):128-38.

131. Subirana I, Fitó M, Diaz O, *et al.* Prediction of coronary disease incidence by biomarkers of inflammation, oxidation, and metabolism. Sci Rep. 2018;8(1):3191.

132. Lin JS, Evans C V., Johnson E, *et al.* Nontraditional Risk Factors in Cardiovascular Disease Risk Assessment: A Systematic Evidence Report for the U.S. Preventive Services

Task Force. Agency for Healthcare Research and Quality (US). 2018. Report No.: 17-05225-EF-1.

133. Wang J, Tan GJ, Han LN, *et al.* Novel biomarkers for cardiovascular risk prediction. J Geriatr Cardiol. 2017;14(2):135-50.

134. Chan D, Ng LL. Biomarkers in acute myocardial infarction. BMC Med. 2010;8:34.

135. Shah PK. Screening asymptomatic subjects for subclinical atherosclerosis: can we, does it matter, and should we? J Am Coll Cardiol. 2010;56(2):98-105.

136. Blaha MJ, Yeboah J, Al Rifai M, *et al.* Providing Evidence for Subclinical CVD in Risk Assessment. Glob Heart. 2016;11(3):275-285.

137. Erbel R, Budoff M. Improvement of cardiovascular risk prediction using coronary imaging: Subclinical atherosclerosis: The memory of lifetime risk factor exposure. Eur Heart J. 2012;33(10):1201-13.

138. Ikonomidis I, Stamatelopoulos K, Lekakis J, *et al.* Inflammatory and non-invasive vascular markers: The multimarker approach for risk stratification in coronary artery disease. Atherosclerosis. 2008;199(1):3-11.

139. Ammirati E, Moroni F, Norata GD, *et al.* Markers of inflammation associated with plaque progression and instability in patients with carotid atherosclerosis. Mediators Inflamm. 2015;2015:718329.

140. Gopal DM, Larson MG, Januzzi JL, *et al.* Biomarkers of cardiovascular stress and subclinical atherosclerosis in the community. Clin Chem. 2014;60(11):1402-08.

141. Chen F-H, Liu T, Xu L, *et al.* Association of Serum Vitamin D Level and Carotid Atherosclerosis: A Systematic Review and Meta-analysis. J Ultrasound Med. 2018;37(6):1293-303.

142. Cinegaglia N, Antoniazzi L, Rosa D, *et al.* Shortening telomere is associated with subclinical atherosclerosis biomarker in omnivorous but not in vegetarian healthy men. Aging. 2019;11(14):5070-80.

143. Greenland P, LaBree L, Azen SP, *et al.* Coronary artery calcium score combined with Framingham score for risk

prediction in asymptomatic individuals. JAMA. 2004;291(2):210-5.

144. Simon A, Chironi G, Levenson J. Comparative performance of subclinical atherosclerosis tests in predicting coronary heart disease in asymptomatic individuals. Eur Heart J. 2007;28(24):2967-71.

145. O'Leary DH, Polak JF, Kronmal RA, *et al.* Carotid-artery intima and media thickness as a risk factor for myocardial infarction and stroke in older adults. Cardiovascular Health Study Collaborative Research Group. N Engl J Med. 1999;340(1):14-22.

146. Polak JF, Pencina MJ, Pencina KM, *et al.* Carotid-wall intima-media thickness and cardiovascular events. N Engl J Med. 2011;365(3):213-21.

147. Nezu T, Hosomi N, Aoki S, Matsumoto M. Carotid Intima-Media Thickness for Atherosclerosis. J Atheroscler Thromb. 2016;23(1):18-31.

148. Costanzo P, Perrone-Filardi P, Vassallo E, *et al.* Does carotid intima-media thickness regression predict reduction of cardiovascular events?: A meta-analysis of 41 randomized trials. J Am Coll Cardiol. 2010;56(24):2006-20.

149. Naqvi TZ, Lee M-S. Carotid intima-media thickness and plaque in cardiovascular risk assessment. JACC Cardiovasc Imaging. 2014;7(10):1025-38.

150. Fowkes FGR, Low LP, Tuta S, *et al.* Ankle-brachial index and extent of atherothrombosis in 8891 patients with or at risk of vascular disease: results of the international AGATHA study. Eur Heart J. 2006;27(15):1861-67.

151. Eldrup N, Sillesen H, Prescott E, Nordestgaard BG. Ankle brachial index, C-reactive protein, and central augmentation index to identify individuals with severe atherosclerosis. Eur Heart J. 2006;27(3):316-22.

152. Alizargar J, Bai CH. Value of the arterial stiffness index and ankle brachial index in subclinical atherosclerosis screening in healthy community-dwelling individuals. BMC Public Health. 2019;19(1):65.

153. Avolio AP, Kuznetsova T, Heyndrickx GR, *et al.* Arterial Flow, Pulse Pressure and Pulse Wave Velocity in Men and Women at Various Ages. Adv Exp Med Biol. 1065:153-68.

154. Kim HL, Kim SH. Pulse Wave Velocity in Atherosclerosis. Front Cardiovasc Med. 2019;6:41.

155. Cavalli-Sforza LL, Bodmer WF. The Genetics of Human Populations. W. H. Freeman. 1971.

156. Karki R, Pandya D, Elston RC, Ferlini C. Defining "mutation" and "polymorphism" in the era of personal genomics. BMC Med Genomics. 2015;8:37.

157. Brookes AJ. The essence of SNPs. Gene. 1999;234(2):177-86.

158. Lander ES, Linton LM, Birren B, *et al.* Initial sequencing and analysis of the human genome. Nature. 2001;409(6822):860-921.

159. Venter JC, Adams MD, Myers EW, *et al.* The Sequence of the Human Genome. Science. 2001;291(5507):1304-51.

160. International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome. Nature. 2004;431(7011):931-45.

161. Consortium TIH. The International HapMap Project. Nature. 2003;426(6968):789-96.

162. Consortium T 1000 GP. A global reference for human genetic variation. Nature. 2015;526(7571):68-74.

163. Tam V, Patel N, Turcotte M, *et al.* Benefits and limitations of genome-wide association studies. Nat Rev Genet. 2019;20(8):467-84.

164. Barrett JC, Cardon LR. Evaluating coverage of genome-wide association studies. Nat Genet. 2006;38(6):659-62.

165. Ozaki K, Ohnishi Y, Iida A, *et al.* Functional SNPs in the lymphotoxin-α gene that are associated with susceptibility to myocardial infarction. Nat Genet. 2002;32(4):650-54.

166. Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature. 2007;447(7145):661-78.

167. Buniello A, MacArthur JAL, Cerezo M, *et al.* The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. Nucleic Acids Res. 2019;47(D1):D1005-D1012.

168. Das S, Abecasis GR, Browning BL. Genotype Imputation from Large Reference Panels. Annu Rev Genomics Hum Genet. 2018;19:73-96.

169. Erdmann J, Kessler T, Munoz Venegas L, Schunkert H. A decade of genome-wide association studies for coronary artery disease: the challenges ahead. Cardiovasc Res. 2018;114(9):1241-57.

170. Elosua R, Sayols-Baixeras S. The Genetics of Ischemic Heart Disease: From Current Knowledge to Clinical Implications. Rev Española Cardiol. 2017:70(9):754-62.

171. Lek M, Karczewski KJ, Minikel EV, *et al.* Analysis of protein-coding genetic variation in 60,706 humans. Nature. 2016;536(7616):285-91.

172. Karczewski KJ, Francioli LC, Tiao G, *et al.* Variation across 141,456 human exomes and genomes reveals the spectrum of loss-of-function intolerance across human protein-coding genes. bioRxiv. 2019:531210.

173. Khera A V, Kathiresan S. Genetics of coronary artery disease: discovery, biology and clinical translation. Nat Rev Genet. 2017;18(6):331-44.

174. Minikel EV, Karczewski KJ, Martin HC, *et al.* Evaluating potential drug targets through human loss-of-function genetic variation. bioRxiv. 2019:530881.

175. Emdin CA, Khera A V., Chaffin M, *et al.* Analysis of predicted loss-of-function variants in UK Biobank identifies variants protective for disease. Nat Commun. 2018;9(1):1613.

176. Dai X, Wiernek S, Evan J, Runge M. Genetics of coronary artery disease and mycardial infarction. World J Cardiol. 2016;8(1):1-23.

177. Helgadottir A, Thorleifsson G, Manolescu A, *et al.* A common variant on chromosome 9p21 affects the risk of myocardial infarction. Science. 2007;316(5830):1491-3.

178. McPherson R, Pertsemlidis A, Kavaslar N, *et al*. A common allele on chromosome 9 associated with coronary heart disease. Science. 2007;316(5830):1488-91.

179. Samani NJ, Erdmann J, Hall AS, *et al*. Genomewide Association Analysis of Coronary Artery Disease. N Engl J Med. 2007;357(5):443-53.

180. CARDIoGRAMplusC4D Consortium, Deloukas P, Kanoni S, Willenborg C, *et al*. Large-scale association analysis identifies new risk loci for coronary artery disease. Nat Genet. 2013;45(1):25-33.

181. Veljkovic N, Zaric B, Djuric I, *et al*. Genetic Markers for Coronary Artery Disease. Medicina. 2018;54(3):E36.

182. Sudlow C, Gallacher J, Allen N, *et al*. UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age. PLoS Med. 2015;12(3):e1001779.

183. Van Der Harst P, Verweij N. Identification of 64 novel genetic loci provides an expanded view on the genetic architecture of coronary artery disease. Circ Res. 2018;122(3):433-43.

184. Roberts R, Marian AJ, Dandona S, Stewart AFR. Genomics in cardiovascular disease. J Am Coll Cardiol. 2013;61(20):2029-37.

185. Kalayinia S, Goodarzynejad H, Maleki M, Mahdieh N. Next generation sequencing applications for cardiovascular disease. Ann Med. 2018;50(2):91-109.

186. Shea J, Agarwala V, Philippakis AA, *et al*. Comparing strategies to fine-map the association of common SNPs at chromosome 9p21 with type 2 diabetes and myocardial infarction. Nat Genet. 2011;43(8):801-5.

187. Bowden DW, An SS, Palmer ND, *et al*. Molecular basis of a linkage peak: exome sequencing and family-based analysis identify a rare genetic variant in the ADIPOQ gene in the IRAS Family Study. Hum Mol Genet. 2010;19(20):4112-20.

188. Musunuru K, Kathiresan S. Genetics of Common, Complex Coronary Artery Disease. Cell. 2019;177(1):132-45.

189. Dainis AM, Ashley EA. Cardiovascular Precision Medicine in the Genomics Era. JACC Basic to Transl Sci. 2018;3(2):313-26.

190. Girelli D, Piubelli C, Martinelli N, *et al.* A decade of progress on the genetic basis of coronary artery disease. Practical insights for the internist. Eur J Intern Med. 2017;41:10-17.

191. Khera AV, Chaffin M, Aragam KG, *et al.* Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. Nat Genet. 2018;50(9):1219-24.

192. Davies NM, Holmes MV, Davey Smith G. Reading Mendelian randomisation studies: a guide, glossary, and checklist for clinicians. BMJ. 2018;362:k601.

193. Holmes MV, Ala-Korpela M, Smith GD. Mendelian randomization in cardiometabolic disease: challenges in evaluating causality. Nat Rev Cardiol. 2017;14(10):577-90.

194. Ference BA, Majeed F, Penumetcha R, Flack JM, Brook RD. Effect of Naturally Random Allocation to Lower Low-Density Lipoprotein Cholesterol on the Risk of Coronary Heart Disease Mediated by Polymorphisms in NPC1L1, HMGCR, or Both: A 2 × 2 Factorial Mendelian Randomization Study. J Am Coll Cardiol. 2015;65(15):1552-61.

195. Do R, Willer CJ, Schmidt EM, *et al.* Common variants associated with plasma triglycerides and risk for coronary artery disease. Nat Genet. 2013;45(11):1345-52.

196. IL6R Genetics Consortium Emerging Risk Factors Collaboration, Sarwar N, Butterworth AS, *et al.* Interleukin-6 receptor pathways in coronary heart disease: a collaborative meta-analysis of 82 studies. Lancet. 2012;379(9822):1205-13.

197. Zacho J, Tybjærg-Hansen A, Jensen JS, *et al.* Genetically elevated C-reactive protein and ischemic vascular disease. N Engl J Med. 2008;359(18):1897-908.

198. Borges MC, Lawlor DA, de Oliveira C, *et al.* Role of adiponectin in coronary heart disease risk: A Mendelian randomization study. Circ Res. 2016;119(3):491-9.

199. Schmidt AF, Swerdlow DI, Holmes M V, *et al*. PCSK9 genetic variants and risk of type 2 diabetes: a mendelian randomisation study. Lancet Diabetes Endocrinol. 2017;5(2):97-105.

200. Visscher PM, Hill WG, Wray NR. Heritability in the genomics era - concepts and misconceptions. Nat Rev Genet. 2008;9(4):255-66.

201. Waddington CH. The Epigenotype. Int J Epidemiol. 2012;41(1):10-3.

202. Nanney DL. Epigenetic control systems. Proc Natl Acad Sci. 1958;44(7):712-7.

203. Jacob F, Monod J. Genetic regulatory mechanisms in the synthesis of proteins. J Mol Biol. 1961;3(3):318-56.

204. Britten RJ, Davidson EH. Gene regulation for higher cells: a theory. Science. 1969;165(3891):349-57.

205. Cavalli G, Heard E. Advances in epigenetics link genetics to the environment and disease. Nature. 2019;571(7766):489-99.

206. Millership SJ, Van de Pette M, Withers DJ. Genomic imprinting and its effects on postnatal growth and adult metabolism. Cell Mol Life Sci. 2019. doi: 10.1007/s00018-019-03197-z.

207. Chao W. Chapter 22: Genomic imprinting. In: Handbook of Epigenetics: The New Molecular and Medical Genetic. Elsevier. 2011:353-79.

208. Brown CJ, Greally JM. A stain upon the silence: genes escaping X inactivation. Trends Genet. 2003;19(8):432-8.

209. Kouzarides T. Chromatin modifications and their function. Cell. 2007;128(4):693-705.

210. Zhou B-R, Bai Y. Chromatin structures condensed by linker histones. Essays Biochem. 2019;63(1):75-87.

211. du Preez LL, Patterton H-G. Secondary structures of the core histone N-terminal tails: their role in regulating chromatin structure. Subcell Biochem. 2013;61:37-55.

212. Pardal AJ, Fernandes-Duarte F, Bowman AJ. The histone chaperoning pathway: from ribosome to nucleosome. Essays Biochem. 2019;63(1):29-43.

213. Li Z, Kono H. Distinct roles of histone H3 and H2A tails in nucleosome stability. Sci Rep. 2016;6:31437.

214. Erler J, Zhang R, Petridis L, *et al*. The role of histone tails in the nucleosome: a computational study. Biophys J. 2014;107(12):2911-22.

215. Talbert PB, Henikoff S. Histone variants on the move: substrates for chromatin dynamics. Nat Rev Mol Cell Biol. 2017;18(2):115-26.

216. Panzeri I, Rossetti G, Pagani M. Chapter 4: Basic principles of noncoding RNAs in epigenetics. In: Medical Epigenetics. 2016:47-63.

217. Wei J-W, Huang K, Yang C, Kang CS. Non-coding RNAs as regulators in epigenetics (Review). Oncol Rep. 2017;37(1):3-9.

218. Esteller M. Non-coding RNAs in human disease. Nat Rev Genet. 2011;12(12):861-74.

219. Collins LJ, Schönfeld B, Chen XS. Chapter 4: The epigenetics of non-coding RNA. In: Handbook of Epigenetics: The New Molecular and Medical. 2011:49-61.

220. Tripathi R, Chakraborty P, Varadwaj PK. Unraveling long non-coding RNAs through analysis of high-throughput RNA-sequencing data. Noncoding RNA Res. 2017;2(2):111-8.

221. Chen L-L. The biogenesis and emerging roles of circular RNAs. Nat Rev Mol Cell Biol. 2016;17(4):205-11.

222. Zakhari S. Alcohol metabolism and epigenetics changes. Alcohol Res Curr Rev. 2013;35(1):6-16.

223. Greenberg MVC, Bourc'his D. The diverse roles of DNA methylation in mammalian development and disease. Nat Rev Mol Cell Biol. 2019. doi: 10.1038/s41580-019-0159-6.

224. Schmitz RJ, Lewis ZA, Goll MG. DNA methylation: shared and divergent features across Eekaryotes. Trends Genet. 2019;S0168-9525(19)30145-3.

225. He Y, Ecker JR. Non-CG methylation in the human genome. Annu Rev Genomics Hum Genet. 2015;16(1):55-77.

226. Zeng Y, Chen T. DNA methylation reprogramming during mammalian development. Genes. 2019;10(4):E257.

227. Kurdyukov S, Bullock M. DNA methylation analysis: choosing the right method. Biology. 2016;5(1):3.

228. Shen L, Waterland RA. Methods of DNA methylation analysis. Curr Opin Clin Nutr Metab Care. 2007;10(5):576-81.

229. Chatterjee A, Rodger EJ, Morison IM, *et al.* Tools and strategies for analysis of genome-wide and gene-specific DNA methylation patterns. Methods in Molecular Biology. 2017;1537:249-277.

230. Illumina. Infinium MethylationEPIC BeadChip Support Resources. Retrieved from: http://emea.support.illumina.com/array/array_kits/infinium-methylationepic-beadchip-kit.html?langsel=/es/. Accessed 14 August 2019.

231. Bibikova M, Barnes B, Tsan C, *et al.* High density DNA methylation array with single CpG site resolution. Genomics. 2011;98(4):288-95.

232. Moran S, Arribas C, Esteller M. Validation of a DNA methylation microarray for 850,000 CpG sites of the human genome enriched in enhancer sequences. Epigenomics. 2016;8(3):389-99.

233. Dedeurwaerder S, Defrance M, Bizet M, *et al.* A comprehensive overview of Infinium HumanMethylation450 data processing. Brief Bioinform. 2014;15(6):929-41.

234. Wilhelm-Benartzi CS, Koestler DC, Karagas MR, *et al.* Review of processing and analysis methods for DNA methylation array data. BrJCancer. 109(6):1394-402.

235. Pidsley R, Y Wong CC, Volta M, *et al.* A data-driven approach to preprocessing Illumina 450K methylation array data. BMC Genomics. 2013;14(1):293.

236. Du P, Zhang X, Huang C-C, *et al.* Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis. BMC Bioinformatics. 2010;11(1):587.

237. Rakyan VK, Down TA, Balding DJ, *et al.* Epigenome-wide association studies for common human diseases. Nat Rev Genet. 2011;12:529-41.

238. Flanagan JM. Epigenome-wide association studies (EWAS): past, present, and future. Methods Mol Biol. 2015;1238:51-63.

239. Michels KB, Binder AM, Dedeurwaerder S, *et al.* Recommendations for the design and analysis of epigenome-wide association studies. Nat Methods. 2013;10(10):949-55.

240. Lin X, Barton S, Holbrook JD. How to make DNA methylome wide association studies more powerful. Epigenomics. 2016;8(8):1117-29.

241. Teschendorff AE, Relton CL. Statistical and integrative system-level analysis of DNA methylation data. Nat Rev Genet. 2018;19(3):129-47.

242. Leenen FAD, Muller CP, Turner JD. DNA methylation: conducting the orchestra from exposure to phenotype? Clin Epigenetics. 2016;8(1):92.

243. Houseman EA, Accomando WP, Koestler DC, *et al.* DNA methylation arrays as surrogate measures of cell mixture distribution. BMC Bioinformatics. 2012;13(1):86.

244. Teschendorff AE, Zheng SC. Cell-type deconvolution in epigenome-wide association studies: a review and recommendations. Epigenomics. 2017;9(5):757-68.

245. Leek JT, Johnson WE, Parker HS, *et al.* The sva package for removing batch effects and other unwanted variation in high-throughput experiments. Bioinformatics. 2012;28(6):882-3.

246. Chen D-P, Lin Y-C, Fann CSJ. Methods for identifying differentially methylated regions for sequence- and array-based data. Brief Funct Genomics. 2016;15(6):485-90.

247. Tobi EW, Goeman JJ, Monajemi R, *et al.* DNA methylation signatures link prenatal famine exposure to growth and metabolism. Nat Commun. 2014;5(1):5592.

248. Tobi EW, Slieker RC, Stein AD, *et al.* Early gestation as the critical time-window for changes in the prenatal environment to affect the adult human blood methylome. Int J Epidemiol. 2015;44(4):1211-23.

249. Andersen GB, Tost J. A summary of the biological processes, disease-associated changes, and clinical applications of DNA methylation. Methods in Molecular Biology. 2018;1708:3-30.

250. Horvath S, Raj K. DNA methylation-based biomarkers and the epigenetic clock theory of ageing. Nat Rev Genet. 2018;19(6):371-84.

251. Gao X, Jia M, Zhang Y, *et al*. DNA methylation changes of whole blood cells in response to active smoking exposure in adults: a systematic review of DNA methylation studies. Clin Epigenetics. 2015;7(1):113.

252. Zhong J, Agha G, Baccarelli AA. The role of DNA methylation in cardiovascular risk and disease. Circ Res. 2016;118(1):119-31.

253. Breitling LP, Salzmann K, Rothenbacher D, *et al*. Smoking, F2RL3 methylation, and prognosis in stable coronary heart disease. Eur Heart J. 2012;33:2841-8.

254. Robertson, KD. DNA methylation and human disease. Nat Rev Genet. 2005;6(8):597-610.

255. Aavik E, Babu M, Ylä-Herttuala S. DNA methylation processes in atheosclerotic plaque. Atherosclerosis. 2019;281:168-79.

256. Khyzha N, Alizada A, Wilson MD, *et al*. Epigenetics of atherosclerosis: emerging mechanisms and methods. Trends Mol Med. 2017;23(4):332-47.

257. Dogan MV., Grumbach IM, Michaelson JJ, Philibert RA. Integrated genetic and epigenetic prediction of coronary heart disease in the Framingham Heart Study. PLoS One. 2018;13(1):e0190549.

258. Ahsan M, Ek WE, Rask-Andersen M, *et al*. The relative contribution of DNA methylation and genetic variants on protein biomarkers for human diseases. PLoS Genet. 2017;13(9):e1007005.

259. Ekelund U, Steene-Johannessen J, Brown WJ, *et al*. Does physical activity attenuate, or even eliminate, the detrimental association of sitting time with mortality? A harmonised meta-analysis of data from more than 1 million men and women. Lancet. 2016;388(10051):1302-10.

260. Elosua R, Redondo A, Segura A, et al. Dose–response association of physical activity with acute myocardial

infarction: Do amount and intensity matter? Prev Med. 2013;57(5):567-72.

261. Lee I-M, Shiroma EJ, Lobelo F, *et al*. Effect of physical inactivity on major non-communicable diseases worldwide: an analysis of burden of disease and life expectancy. Lancet. 2012;380(9838):219-29.

262. Hawley JA, Hargreaves M, Joyner MJ, *et al*. Integrative biology of exercise. Cell. 2014;159(4):738-49.

263. Voisin S, Eynon N, Yan X, Bishop DJ. Exercise training and DNA methylation in humans. Acta Physiol. 2015;213(1):39-59.

264. Grazioli E, Dimauro I, Mercatelli N, *et al*. Physical activity in the prevention of human diseases: role of epigenetic modifications. BMC Genomics. 2017;18(8):802.

265. Denham J, O'Brien BJ, Marques FZ, *et al*. Changes in the leukocyte methylome and its effect on cardiovascular-related genes after exercise. J Appl Physiol. 2015;118(4):475-88.

266. Jacques M, Hiam D, Craig J, *et al*. Epigenetic changes in healthy human skeletal muscle following exercise–a systematic review. Epigenetics. 2019;14(7):633-48.

267. Rönn T, Volkov P, Davegårdh C, *et al*. A Six months exercise intervention influences the genome-wide DNA methylation pattern in human adipose tissue. PLoS Genet. 2013;9(6):e1003572.

268. Fabre O, Ingerslev LR, Garde C, *et al*. Exercise training alters the genomic response to acute exercise in human adipose tissue. Epigenomics. 2018;10(8):1033-50.

269. Sillanpää E, Ollikainen M, Kaprio J, *et al*. Leisure-time physical activity and DNA methylation age-a twin study. Clin Epigenetics. 2019;11(1):12.

270. Sailani MR, Halling JF, Møller HD, *et al*. Lifelong physical activity is associated with promoter hypomethylation of genes involved in metabolism, myogenesis, contractile properties and oxidative stress resistance in aged human skeletal muscle. Sci Rep. 2019;9(1):3272.

271. Schenk A, Koliamitra C, Bauer CJ, *et al*. Impact of acute aerobic exercise on genome-wide DNA-methylation in Natural Killer Cells-A Pilot study. Genes. 2019;10(5):E380.

272. van Roekel EH, Dugué P-A, Jung C-H, *et al*. Physical activity, television viewing time, and DNA methylation in peripheral blood. Med Sci Sport Exerc. 2019;51(3):490-8.

273. Ekelund U, Tarp J, Steene-Johannessen J, *et al*. Dose-response associations between accelerometry measured physical activity and sedentary time and all cause mortality: systematic review and harmonised meta-analysis. BMJ. 2019;366:I4570.

274. Cárdenas-Fuentes G, Subirana I, Martinez-Gonzalez MA, *et al*. Multiple approaches to associations of physical activity and adherence to the Mediterranean diet with all-cause mortality in older adults: the PREvención con DIeta MEDiterránea study. Eur J Nutr. 2019;58(4):1569-78.

275. Fang X, Han D, Cheng Q, *et al*. Association of levels of physical activity with risk of Parkinson disease. JAMA Netw Open. 2018;1(5):e182421.

276. Li T, Wei S, Shi Y, *et al*. The dose-response effect of physical activity on cancer mortality: findings from 71 prospective cohort studies. Br J Sports Med. 2016;50(6):339-45.

277. Aune D, Norat T, Leitzmann M, *et al*. Physical activity and the risk of type 2 diabetes: a systematic review and dose–response meta-analysis. Eur J Epidemiol. 2015;30(7):529-42.

278. Joehanes R, Just AC, Marioni RE, *et al*. Epigenetic signatures of cigarette smoking. circ cardiovasc genet. 2016;9(5):436-47.

279. Imgenberg-Kreuz J, Carlsson Almlöf J, Leonard D, *et al*. DNA methylation mapping identifies gene regulatory effects in patients with systemic lupus erythematosus. Ann Rheum Dis. 2018;77(5):736-43.

280. Gross AM, Jaeger PA, Kreisberg JF, *et al*. Methylome-wide analysis of chronic HIV infection reveals five-year increase in biological age and epigenetic targeting of HLA. Mol Cell. 2016;62(2):157-68.

281. de Vega WC, Erdman L, Vernon SD, *et al*. Integration of DNA methylation & health scores identifies subtypes in myalgic encephalomyelitis/chronic fatigue syndrome. Epigenomics. 2018;10(5):539-57.

282. Kulminski AM, Huang J, Loika Y, *et al.* Strong impact of natural-selection–free heterogeneity in genetics of age-related phenotypes. Aging. 2018;10(3):492-514.

283. Kichaev G, Bhatia G, Loh PR, *et al.* Leveraging polygenic functional enrichment to improve GWAS power. Am J Hum Genet. 2019;104(1):65-75.

284. Bhatt-Wessel B, Jordan TW, Miller JH, *et al.* Role of DGAT enzymes in triacylglycerol metabolism. Arch Biochem Biophys. 2018;655:1-11.

285. Wang Y, Xu D. Effects of aerobic exercise on lipids and lipoproteins. Lipids Health Dis. 2017;16(1):132.

286. Lee T-W, Bai K-J, Lee T-I, *et al.* PPARs modulate cardiac metabolism and mitochondrial function in diabetes. J Biomed Sci. 2017;24(1):5.

287. Herzig S, Shaw RJ. AMPK: guardian of metabolism and mitochondrial homeostasis. Nat Rev Mol Cell Biol. 2017;19(2):121-35.

288. King-Himmelreich TS, Schramm S, Wolters MC, *et al.* The impact of endurance exercise on global and AMPK gene-specific DNA methylation. Biochem Biophys Res Commun. 2016;474(2):284-90.

289. Yin Y, Morgunova E, Jolma A, *et al.* Impact of cytosine methylation on DNA binding specificities of human transcription factors. Science. 2017;356(6337):eaaj2239.

290. Roadmap Epigenomics Consortium, Kundaje A, Meuleman W, *et al.* Integrative analysis of 111 reference human epigenomes. Nature. 2015;518(7539):317-29.

291. Polli A, Ickmans K, Godderis L, *et al.* When environment meets genetics: a clinical review of the epigenetics of pain, psychological factors, and physical activity. Arch Phys Med Rehabil. 2019;100(6):1153-61.

292. Rönn T, Ling C. Effect of exercise on DNA methylation and metabolism in human adipose tissue and skeletal muscle. Epigenomics. 2013;5(6):603-05.

293. Ross R, Goodpaster BH, Koch LG, *et al.* Precision exercise medicine: understanding exercise response variability. Br J Sports Med. 2019;53(18):1141-53.

294. Yusuf S, Hawken S, Ôunpuu S, *et al.* Effect of potentially modifiable risk factors associated with myocardial infarction in 52 countries (the INTERHEART study): case-control study. Lancet. 2004;364(9438):937-52.

295. Sayols-Baixeras S. DNA methylation and its relationship with lifestyle, environmental and cardiovascular risk factors. Universitat Pompeu Fabra; 2019.

296. Sayols-Baixeras S, Lluís-Ganella C, Subirana I, *et al.* Identification of a new locus and validation of previously reported loci showing differential methylation associated with smoking. The REGICOR study. Epigenetics. 2015;10(12):1156-65.

297. Sayols-Baixeras S, Subirana I, Lluis-Ganella C, *et al.* Identification and validation of seven new loci showing differential DNA methylation related to serum lipid profile: an epigenome-wide approach. The REGICOR study. Hum Mol Genet. 2016;25(20):ddw285.

298. Sayols-Baixeras S, Subirana I, Fernández-Sanlés A, *et al.* DNA methylation and obesity traits: An epigenome-wide association study. The REGICOR study. Epigenetics. 2017;12(10): 909-16.

299. Roetker NS, Pankow JS, Bressler J, *et al.* Prospective Study of Epigenetic Age Acceleration and Incidence of Cardiovascular Disease Outcomes in the ARIC Study (Atherosclerosis Risk in Communities). Circ Genomic Precis Med. 2018;11(3):e001937.

300. Breton C V., Park C, Siegmund K, *et al.* NOS1 Methylation and Carotid Artery Intima-Media Thickness in Children. Circ Cardiovasc Genet. 2014;7(2):116-22.

301. Murray R, Bryant J, Titcombe P, *et al.* DNA methylation at birth within the promoter of ANRIL predicts markers of cardiovascular risk at 9 years. Clin Epigenetics. 2016;8(1):90.

302. Istas G, Declerck K, Pudenz M, *et al.* Identification of differentially methylated BRCA1 and CRISP2 DNA regions as blood surrogate markers for cardiovascular disease. Sci Rep. 2017;7(1):5120.

303. Petersen A-K, Zeilinger S, Kastenmüller G, *et al.* Epigenetics meets metabolomics: an epigenome-wide association study

with blood serum metabolic traits. Hum Mol Genet. 2014;23(2):534-45.

304. Perrucci GL, Gowran A, Zanobini M, *et al.* Peptidyl-prolyl isomerases: a full cast of critical actors in cardiovascular diseases. Cardiovasc Res. 2015;106(3):353-64.

305. Philibert RA, Penaluna B, White T, *et al.* A pilot examination of the genome-wide DNA methylation signatures of subjects entering and exiting short-term alcohol dependence treatment programs. Epigenetics. 2014;9(9):1212-9.

306. Liu C, Marioni RE, Hedman ÅK, et al. A DNA methylation biomarker of alcohol consumption. Mol Psychiatry. 2018;23(2):422-33.

307. El Kasmi KC, Smith AM, Williams L, *et al.* Cutting Edge: A Transcriptional Repressor and Corepressor Induced by the STAT3-Regulated Anti-Inflammatory Signaling Pathway. J Immunol. 2007;179(11):7215-9.

308. Rönn T, Volkov P, Gillberg L, *et al.* Impact of age, BMI and HbA1c levels on the genome-wide DNA methylation and mRNA expression patterns in human adipose tissue and identification of epigenetic biomarkers in blood. Hum Mol Genet. 2015;24(13):3792-813.

309. Wahl S, Drong A, Lehne B, *et al.* Epigenome-wide association study of body mass index, and the adverse outcomes of adiposity. Nature. 2016;541(7635):81-6.

310. Ligthart S, Marzi C, Aslibekyan S, *et al.* DNA methylation signatures of chronic low-grade inflammation are associated with complex diseases. Genome Biol. 2016;17(1):255.

311. Ek WE, Hedman AK, Enroth S, *et al.* Genome-wide DNA methylation study identifies genes associated with the cardiovascular biomarker GDF-15. Hum Mol Genet. 2016;25:817-27.

312. Dertinger SD, Silverstone AE, Gasiewicz TA. Influence of aromatic hydrocarbon receptor-mediated events on the genotoxicity of cigarette smoke condensate. Carcinogenesis. 1998;19(11):2037-42.

313. Vogel CFA, Haarmann-Stemmann T. The aryl hydrocarbon receptor repressor - More than a simple feedback inhibitor of

AhR signaling: Clues for its role in inflammation and cancer. Curr Opin Toxicol. 2017;2:109-119.

314. Rothhammer V, Quintana FJ. The aryl hydrocarbon receptor: an environmental sensor integrating immune responses in health and disease. Nat Rev Immunol. 2019;19(3):184-97.

315. Philibert R, Hollenbeck N, Andersen E, *et al.* A quantitative epigenetic approach for the assessment of cigarette consumption. Front Psychol. 2015;6:656.

316. Philibert R, Dogan M, Noel A, *et al.* Dose Response and Prediction Characteristics of a Methylation Sensitive Digital PCR Assay for Cigarette Consumption in Adults. Front Genet. 2018;9:137.

317. Joubert BR, Håberg SE, Nilsen RM, *et al.* 450K epigenome-wide scan identifies differential DNA methylation in newborns related to maternal smoking during pregnancy. Environ Health Perspect. 2012;120(10):1425-31.

318. Bojesen SE, Timpson N, Relton C, *et al.* AHRR (cg05575921) hypomethylation marks smoking behaviour, morbidity and mortality. Thorax. 2017;72(7):646-53.

319. Bakulski KM, Dou J, Lin N, *et al.* DNA methylation signature of smoking in lung cancer is enriched for exposure signatures in newborn and adult blood. Sci Rep. 2019;9(1):4576.

320. Kodal JB, Kobylecki CJ, Vedel-Krogh S, *et al.* AHRR hypomethylation, lung function, lung function decline and respiratory symptoms. Eur Respir J. 2018;51(3):1701512.

321. Andersen AM, Ryan PT, Gibbons FX, *et al.* A Droplet Digital PCR Assay for Smoking Predicts All-Cause Mortality. J Insur Med. 2018;47(4):220-9.

322. Casals N, Zammit V, Herrero L, *et al.* Carnitine palmitoyltransferase 1C: From cognition to cancer. Prog Lipid Res. 2016;61:134-48.

323. Das M, Sha J, Hidalgo B, *et al.* Association of DNA Methylation at CPT1A Locus with Metabolic Syndrome in the Genetics of Lipid Lowering Drugs and Diet Network (GOLDN) Study. PLoS One. 2016;11(1):e0145789.

324. Kriebel J, Herder C, Rathmann W, *et al.* Association between DNA Methylation in Whole Blood and Measures of Glucose

Metabolism: KORA F4 Study. PLoS One. 2016;11(3):e0152314.

325. Kulkarni H, Kos MZ, Neary J, *et al.* Novel epigenetic determinants of type 2 diabetes in Mexican-American families. Hum Mol Genet. 2015;24(18):5330-44.

326. Aslibekyan S, Demerath EW, Mendelson M, *et al.* Epigenome-wide study identifies novel methylation loci associated with body mass index and waist circumference. Obesity (Silver Spring). 2015;23(7):1493-501.

327. 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. Liming Liang Devin M Absher Hum Mol Genet. 2015;24(15):4464-4479.

328. Petersen PS, Lei X, Wolf RM, *et al.* CTRP7 deletion attenuates obesity-linked glucose intolerance, adipose tissue inflammation, and hepatic stress. Am J Physiol Metab. 2017;312(4):E309-25.

329. Dekkers KF, van Iterson M, Slieker RC, *et al.* Blood lipids influence DNA methylation in circulating cells. Genome Biol. 2016;17(1):138.

330. Sayols-Baixeras S, Tiwari HK, Aslibekyan SW. Disentangling associations between DNA methylation and blood lipids: a Mendelian randomization approach. BMC Proc. 2018;12(S9):23.

331. Sayols-Baixeras S, Irvin MR, Elosua R, *et al.* Epigenetics of Lipid Phenotypes. Curr Cardiovasc Risk Rep. 2016;10(10):31.

332. Wan M, Bennett BD, Pittman GS, *et al.* Identification of Smoking-Associated Differentially Methylated Regions Using Reduced Representation Bisulfite Sequencing and Cell type–Specific Enhancer Activation and Gene Expression. Environ Health Perspect. 2018;126(4):47015.

333. Tsai P-C, Glastonbury CA, Eliot MN, *et al.* Smoking induces coordinated DNA methylation and gene expression changes in adipose tissue with consequences for metabolic health. Clin Epigenetics. 2018;10(1):126.

334. Shenker NS, Polidoro S, van Veldhoven K, *et al.* Epigenome-wide association study in the European Prospective Investigation into Cancer and Nutrition (EPIC-Turin) identifies novel genetic loci associated with smoking. Hum Mol Genet. 2013;22(5):843-51.

335. Monick MM, Beach SRH, Plume J, *et al.* Coordinated Changes in AHRR Methylation in Lymphoblasts and Pulmonary Macrophages from Smokers. Am J Med Genet B Neuropsychiatr Genet. 2012;159B(2):141-51.

336. Reynolds LM, Wan M, Ding J, *et al.* DNA Methylation of the Aryl Hydrocarbon Receptor Repressor Associations With Cigarette Smoking and Subclinical Atherosclerosis. Circ Cardiovasc Genet. 2015;8(5):707-16.

337. Arimilli S, Madahian B, Chen P, *et al.* Gene expression profiles associated with cigarette smoking and moist snuff consumption. BMC Genomics. 2017;18(1):156.

338. Chen Y, Widschwendter M, Teschendorff AE. Systems-epigenomics inference of transcription factor activity implicates aryl-hydrocarbon-receptor inactivation as a key event in lung cancer development. Genome Biol. 2017;18(1):236.

339. Ma B, Wilker EH, Willis-Owen SAG, *et al.* Predicting DNA methylation level across human tissues. Nucleic Acids Res. 2014;42(6):3515-28.

340. Baccarelli A, Wright R, Bollati V, *et al.* Ischemic heart disease and stroke in relation to blood DNA methylation. Epidemiology. 2010;21:819-28.

341. Guarrera S, Fiorito G, Onland-Moret NC, *et al.* Gene-specific DNA methylation profiles and LINE-1 hypomethylation are associated with myocardial infarction risk. Clin Epigenetics. 2015;7:133.

342. Nakatochi M, Ichihara S, Yamamoto K, *et al.* Epigenome-wide association of myocardial infarction with DNA methylation sites at loci related to cardiovascular disease. Clin Epigenetics. 2017;9(1):54.

343. Wang X, Liu A-H, Jia Z-W, *et al.* Genome-wide DNA methylation patterns in coronary heart disease. Herz. 2018;43(7):656-62.

344. Pepin ME, Ha C-M, Crossman DK, *et al.* Genome-wide DNA methylation encodes cardiac transcriptional reprogramming in human ischemic heart failure. Lab Invest. 2019;99(3):371-86.

345. Li B, Feng Z-H, Sun H, *et al.* The blood genome-wide DNA methylation analysis reveals novel epigenetic changes in human heart failure. Eur Rev Med Pharmacol Sci. 2017;21(8):1828-36.

346. Thunders M, Holley A, Harding S, *et al.* Using NGS-methylation profiling to understand the molecular pathogenesis of young MI patients who have subsequent cardiac events. Epigenetics. 2019;14(6):536-44.

347. Hedman ÅK, Mendelson MM, Marioni RE, *et al.* Epigenetic patterns in blood associated with lipid traits predict incident coronary heart disease events and are enriched for results from genome-wide association studies. Circ Cardiovasc Genet. 2017;10(1):e001487.

348. Aslibekyan S, Agha G, Colicino E, *et al.* Association of methylation signals with incident coronary heart disease in an epigenome-wide assessment of circulating tumor necrosis factor α. JAMA Cardiol. 2018;3(6):463-72.

349. Westerman K, Sebastiani P, Jacques P, *et al.* DNA methylation modules associate with incident cardiovascular disease and cumulative risk factor exposure. bioRxiv. 2018:471722.

350. Ward-Caviness CK, Agha G, Chen BH, *et al.* Analysis of repeated leukocyte DNA methylation assessments reveals persistent epigenetic alterations after an incident myocardial infarction. Clin Epigenetics. 2018;10(1):161.

351. Agha G, Mendelson MM, Ward-Caviness CK, *et al.* Blood Leukocyte DNA Methylation Predicts Risk of Future Myocardial Infarction and Coronary Heart Disease. Circulation. 2019;140(8):645-57.

352. Li G, Gu H-M, Zhang D-W. ATP-binding cassette transporters and cholesterol translocation. IUBMB Life. 2013;65(6):505-12.

353. Olivier M, Tanck MW, Out R, *et al.* Human ATP–Binding Cassette G1 Controls Macrophage Lipoprotein Lipase Bioavailability and Promotes Foam Cell Formation. Arterioscler Thromb Vasc Biol. 2012;32(9):2223-31.

354. Nafisinia M, Riley LG, Gold WA, *et al.* Compound heterozygous mutations in glycyl-tRNA synthetase (GARS) cause mitochondrial respiratory chain dysfunction. PLoS One. 2017;12(6):e0178125.

355. Svane A, Soerensen M, Lund J, *et al.* DNA Methylation and all-cause mortality in middle-aged and elderly danish twins. Genes (Basel). 2018;9(2):78.

356. Zhang Y, Yang R, Burwinkel B, *et al.* F2RL3 methylation in blood DNA is a strong predictor of mortality. Int J Epidemiol. 2014;43:1215-25.

357. Petersen PS, Lei X, Wolf RM, *et al.* CTRP7 deletion attenuates obesity-linked glucose intolerance, adipose tissue inflammation, and hepatic stress. Am J Physiol Endocrinol Metab. 2017;312(4):E309-25.

358. Dowal L, Yang W, Freeman MR, *et al.* Proteomic analysis of palmitoylated platelet proteins. Blood. 2011;118(13):e62-73.

359. Ronimus RS, Morgan HW. Cloning and biochemical characterization of a novel mouse ADP-dependent glucokinase. Biochem Biophys Res Commun. 2004;315(3):652-658.

360. Johnson MP, Keyho R, Blackburn NB, *et al.* Glycated serum protein genetics and pleiotropy with cardiometabolic risk factors. J Diabetes Res. 2019;2019:2310235.

361. Zhang Y, Mao D, Roswit WT, *et al.* PARP9-DTX3L ubiquitin ligase targets host histone H2BJ and viral 3C protease to enhance interferon signaling and control viral infection. Nat Immunol. 2015;16(12):1215-27.

362. Iwata H, Goettsch C, Sharma A, *et al.* PARP9 and PARP14 cross-regulate macrophage activation via STAT1 ADP-ribosylation. Nat Commun. 2016;7(1):12849.

363. Peiró C, Romacho T, Azcutia V, *et al.* Inflammation, glucose, and vascular cell damage: the role of the pentose phosphate pathway. Cardiovasc Diabetol. 2016;15(1):82.

364. Matsui R, Xu S, Maitland KA, *et al.* Glucose-6-phosphate dehydrogenase deficiency decreases vascular superoxide and atherosclerotic lesions in apolipoprotein E(-/-) mice. Arterioscler Thromb Vasc Biol. 2006;26(4):910-6.

365. Beutler E, Mitchell M, Fornera S, *et al.* Special modifications of the fluorescent screening method for glucose-6-phosphate dehydrogenase deficiency. Blood. 1968;32(5):816-8.

366. Zemmour H, Planer D, Magenheim J, *et al.* Non-invasive detection of human cardiomyocyte death using methylation patterns of circulating DNA. Nat Commun. 2018;9(1):1443.

367. Lluis-Ganella C, Subirana I, Lucas G, *et al.* Assessment of the value of a genetic risk score in improving the estimation of coronary risk. Atherosclerosis. 2012;222(2):456-63.

368. Levine ME, Lu AT, Quach A, *et al.* An epigenetic biomarker of aging for lifespan and healthspan. Aging. 2018;10(4):573-91.

369. Lu AT, Quach A, Wilson JG, *et al.* DNA methylation GrimAge strongly predicts lifespan and healthspan. Aging. 2019;11(2):303-27.

370. Gaunt TR, Shihab HA, Hemani G, *et al.* Systematic identification of genetic influences on methylation across the human life course. Genome Biol. 2016;17(1):61.

371. Heiss JA, Brenner H. Impact of confounding by leukocyte composition on associations of leukocyte DNA methylation with common risk factors. Epigenomics. 2017;9(5):659-68.

372. Goh WWB, Wang W, Wong L. Why batch effects matter in omics data, and how to avoid them. Trends Biotechnol. 2017;35(6):498-507.

373. Horvath S. DNA methylation age of human tissues and cell types. Genome Biol. 2013;14(10):R115.

374. Hannum G, Guinney J, Zhao L, *et al.* Genome-wide methylation profiles reveal quantitative views of human aging rates. Mol Cell. 2013;49(2):359-67.

375. Horvath S, Gurven M, Levine ME, *et al.* An epigenetic clock analysis of race/ethnicity, sex, and coronary heart disease. Genome Biol. 2016;17(1):171.

376. Perna L, Zhang Y, Mons U, *et al.* Epigenetic age acceleration predicts cancer, cardiovascular, and all-cause mortality in a German case cohort. Clin Epigenetics. 2016;8:64.

377. Schildgen V, Schildgen O. How is a molecular polymorphism defined? Cancer. 2013;119(9):1608.

378. Salzberg SL. Open questions: How many genes do we have? BMC Biol. 2018;16(1):94

379. Gardiner-Garden M, Frommer M. CpG Islands in vertebrate genomes. J Mol Biol. 1987;196(2):261-82.

380. Wu H, Caffo B, Jaffee HA, *et al*. Redefining CpG islands using hidden Markov models. Biostatistics. 2010;11(3):499.

"If you obey all the rules, you miss all the fun."

Katharine Hepburn (1907-2003).

Pepe was one of my grandfathers, whom I never met. He was a peasant, and I have been told he was a really intelligent man and a book-devourer. He was taken miles away from home to fight in the Spanish Civil War, but he ended up practicing as a physician's assistant on the front line of combat because he was diagnosed with a heart murmur. Once back in Xuño, our hometown, he continued treating patients. What a coincidence that heart diseases have shaped both his career and mine!