

Focus on: Genetics, Epigenetics, and Cardiovascular Disease (I)

Basic Concepts in Molecular Biology Related to Genetics and Epigenetics

Dolores Corella^{a,b,*} and Jose M. Ordovas^{c,d,e}^aDepartamento de Medicina Preventiva y Salud Pública, Facultad de Medicina, Universidad de Valencia, Valencia, Spain^bCIBER Fisiopatología de la Obesidad y Nutrición, Instituto de Salud Carlos III, Madrid, Spain^cDepartamento de Epidemiología Cardiovascular y Genética de Poblaciones, Centro Nacional de Investigaciones Cardiovasculares (CNIC), Madrid, Spain^dInstituto Madrileño de Estudios Avanzados (IMDEA) Alimentación, Madrid, Spain^eNutrition and Genomics Laboratory, Jean Mayer USDA Human Nutrition Research Center on Aging, Tufts University, Boston, Massachusetts, United States

Article history:

Available online 13 June 2017

Keywords:

Genetics

Epigenetics

Polymorphism

Methylation

ABSTRACT

The observation that “one size does not fit all” for the prevention and treatment of cardiovascular disease, among other diseases, has driven the concept of precision medicine. The goal of precision medicine is to provide the best-targeted interventions tailored to an individual's genome. The human genome is composed of billions of sequence arrangements containing a code that controls how genes are expressed. This code depends on other nonstatic regulators that surround the DNA and constitute the epigenome. Moreover, environmental factors also play an important role in this complex regulation. This review provides a general perspective on the basic concepts of molecular biology related to genetics and epigenetics and a glossary of key terms. Several examples are given of polymorphisms and genetic risk scores related to cardiovascular risk. Likewise, an overview is presented of the main epigenetic regulators, including DNA methylation, methylcytosine-phosphate-guanine-binding proteins, histone modifications, other histone regulations, micro-RNA effects, and additional emerging regulators. One of the greatest challenges is to understand how environmental factors (diet, physical activity, smoking, etc.) could alter the epigenome, resulting in healthy or unhealthy cardiovascular phenotypes. We discuss some gene-environment interactions and provide a methodological overview.

© 2017 Sociedad Española de Cardiología. Published by Elsevier España, S.L.U. All rights reserved.

Conceptos básicos en biología molecular relacionados con la genética y la epigenética

RESUMEN

La observación de que «lo mismo no sirve para todos» en la prevención y el tratamiento de las enfermedades cardiovasculares, entre otras, ha propulsado el concepto de medicina de precisión. Su objetivo es proporcionar las mejores intervenciones basadas en la información adicional que aporta el genoma. El genoma humano se compone de miles de millones de pares de bases que contienen un código que controla cómo se expresan los genes. Este código depende de reguladores no estáticos que rodean el ADN y constituyen el epigenoma. Además, los factores ambientales también desempeñan un papel importante en esta compleja regulación. Se presenta una perspectiva general sobre los conceptos básicos de la biología molecular relacionados con la genética y la epigenética y un glosario de términos clave, se revisan varios ejemplos de polimorfismos y escalas de riesgo genético relacionadas con el riesgo cardiovascular, y se proporciona una visión general de los principales reguladores epigenéticos, como la metilación del ADN, las proteínas de unión a metilcitosina-fosfato-guanina, las modificaciones de histonas, otras regulaciones de histonas, los efectos de los microARN y otros reguladores emergentes. Otro desafío es entender cómo los factores ambientales (dieta, ejercicio, tabaco, etc.) pueden alterar el epigenoma y resultar en fenotipos saludables o no. Se comentan algunas interacciones entre gen y ambiente y se proporciona una visión metodológica general.

© 2017 Sociedad Española de Cardiología. Publicado por Elsevier España, S.L.U. Todos los derechos reservados.

Palabras clave:

Genética

Epigenética

Polimorfismo

Metilación

SEE RELATED CONTENT:

<http://dx.doi.org/10.1016/j.rec.2017.03.012>, Rev Esp Cardiol. 2017;70:696–698.

* Corresponding author: Unidad de Epidemiología Genética y Molecular, Universidad de Valencia, Blasco Ibáñez 15, 46010 Valencia, Spain.

E-mail address: dolores.corella@uv.es (D. Corella).<http://dx.doi.org/10.1016/j.rec.2017.05.011>

1885-5857/© 2017 Sociedad Española de Cardiología. Published by Elsevier España, S.L.U. All rights reserved.

Abbreviations

CpG: cytosine-phosphate-guanine
 GRS: genetic risk score
 GWAS: genome-wide association study
 SNP: single-nucleotide polymorphism

INTRODUCTION

Technological advances made during the Human Genome Project and after its completion have greatly reduced costs and increased the immediacy of access to genetic test results.¹ Genetic markers can now be determined as part of routine investigation of large patient cohorts^{2,3} and clinical trials.^{4,5} The fruits of decades of research into genetic markers have deepened our understanding of the molecular bases of the most common diseases, including cardiovascular diseases and their intermediate phenotypes.^{6,7} However, the huge amount of information on new genes linked to the various diseases and their possible environmental modulation remains in the theoretical realm and subsequent diagnostic, preventive, and therapeutic validation is required before these markers can be translated to clinical practice and applied to the general population. To accelerate this process, an initiative called precision medicine was launched in 2015 in the United States.⁸ As detailed by Collins and Varmus in their 2015 publication on this approach,⁸ the concept of precision medicine—defined as “the implementation of prevention and treatment strategies that take individual variability into account to optimize results”—is not new and has been used in one form or another for many years, such as when patients are blood typed to guide blood transfusions. However, the novelty lies in the huge number of new genetic markers that can help to improve our understanding of patients' genetic risk and predict interindividual treatment responses. In addition to genetic markers based on DNA sequences, numerous other -omic biomarkers are available (eg, epigenomic, transcriptomic, metabolomic, proteomic). In conjunction with bioinformatics and new computational tools for the handling and integration of these data, these biomarkers can provide highly valuable information that markedly improves disease prevention and treatment. Although initial efforts in precision medicine have largely focused on cancer,⁸ advances are also being made in the field of cardiovascular disease that will allow goals to be reached in this new era of medicine.⁹

For precision medicine to move from dream to reality,¹⁰ sustained research effort is needed to incorporate information gleaned on -omic markers into clinical trials and other epidemiological studies and generate results with a high level of scientific evidence to guide decisions in the new era.⁸ This step requires medical and other biomedical scientific professionals to acquire a solid knowledge base in -omics in order to better critically interpret and confront the new challenges. Although the central dogma of biology, proposed by Francis Crick in 1958, proposed a unidirectional flow of the transmission and expression of genetic information, namely, transcription of DNA into messenger RNA and translation of messenger RNA into protein, the factor that finally performs the cellular action,¹¹ it is now known that this pathway is not universal and that powerful regulatory elements can allow the same DNA to give rise to 2 or 3 distinct proteins. Greater understanding is thus required of not only the genetic elements, but also the epigenetic ones contributing to this regulation. Accordingly, the aim of this work is to provide a succinct and up-to-date overview of the basic molecular biology concepts related to genetics and

epigenetics. Our approach involves various examples of the studies performed, focusing on cardiovascular diseases, and culminating in a methodological reflection on what are known as gene-environment interactions. Other medical journals such as the *New England Journal of Medicine*^{12,13} and *Journal of the American Medical Association*,^{14–16} as well as more specialized publications,^{17–20} have also published reviews of the basic concepts in genetics and epigenetics and should be consulted to complement this work.

THE HUMAN GENOME

The human genome contains about 6 billion DNA base pairs (adenine, thymine, guanine, and cytosine) and is organized into 23 pairs of chromosomes. There are an estimated 20 000 to 25 000 genes in the human genome, much fewer than originally predicted,¹² partly because the same gene can give rise to various proteins,¹³ as described below. Because each DNA base pair is about 0.34 nm in length, each diploid cell would contain about 2 meters of DNA if it were fully uncoiled. In total, the human body would contain 100 billion meters of DNA.²¹ Histones are in charge of compacting the DNA so that it can fit within the microscopic cell nucleus. Histones are a family of small positively charged proteins named H1, H2A, H2B, H3, and H4.²² Because DNA is negatively charged, due to the phosphate groups in the phosphate-sugar backbone, histones strongly bind to DNA. The basic structural and functional unit of chromatin is the nucleosome, which contains 8 histone proteins and about 146 DNA base pairs. Nucleosomes in turn form part of another structure called the chromatosome, with each chromatosome packaging an average 100 million base pairs. Thus, each chromosome is a long chain of nucleosomes.²¹

Chromatin is in turn classified into euchromatin and heterochromatin. These 2 forms show differences in staining, structure, and function but euchromatin is basically a more relaxed and transcriptionally active structure that encompasses most genes, whereas heterochromatin is denser and contains more repeat sequences, such as those found in telomeres. The latter are highly repetitive regions located at the end of chromosomes. The most repeated sequence in human telomeres is 5'-TTAGGG-3', with sometimes even more than 2000 repeats. There is a complex of 6 proteins associated with telomeres called shelterin or telomere protection complex, composed of TRF1 and TRF2, that in turn interacts with RAP1, TPP1, POT1, and TIN2 to associate with telomeric DNA.²² The enzyme in charge of telomere extension is called telomerase. This reverse transcriptase maintains telomere length; using RNA as a template and in conjunction with specific accessory proteins, it repeatedly adds the sequence d(TTAGGG) at the extreme 3' end of telomeric DNA.²² Various studies have associated telomere length with different diseases.²³ It is generally believed that shorter telomeres are linked to greater aging and higher cardiovascular risk.²⁴

Human Genome Sequence Variations

Genes constitute the transcriptionally active part of chromosomes. The basic structure of a gene is divided into introns and exons.¹⁹ Introns are noncoding sequences, whereas exons are coding. There are also noncoding regions at the start and end of each gene called the 5'-untranslated region (UTR) and 3'-UTR. The initial region contains the gene promoter and the end region is highly involved in regulation by microRNA. The DNA sequences can be subject to variations of different types (eg, insertions, deletions, repeat expansions).^{14,19} The best known are single nucleotide changes, more commonly known by the abbreviation

SNP (single nucleotide polymorphism).¹⁴ These SNPs can be found in all gene regions. Polymorphisms found in introns were once thought to have no effect on the amino acid sequence of the resulting protein and were said to be nonfunctional. However, the significance of polymorphisms in noncoding regions is now better understood: such variations can actually affect gene functionality without the involvement of an amino acid change, with or without the binding of other regulators, as discussed below. Polymorphisms found in exons do not always cause amino acid changes because, even if there is a base change, the corresponding triplet (group of 3 nucleotides determining an amino acid) that is generated can still code for the same amino acid due to the “degenerate” nature of the genetic code. Accordingly, if the combinations of the 4 elements (4 DNA bases) and 3 bases per codon are counted, they would code for 64 distinct amino acids. However, there are only 20 different amino acids because, as observed by Wittmann in 1962, an amino acid can be encoded by more than 1 codon.

In quantitative terms, 2 unrelated people are estimated to share more than 99% of DNA sequences but, taking into account the billions of base pairs forming the genome, the DNA sequences of 2 unrelated people can vary by over 20 million bases.^{12,13} Some of these base changes can increase the risk of disease, as happens in the types of disorder known as monogenic diseases. These conditions show high penetrance, with a single base change in a specific DNA region sufficient to give rise to a well-characterized disease. Various types of phenotypes have been linked to

cardiovascular diseases with a monogenic origin, such as monogenic familial hypercholesterolemia caused by one or more mutations in the low-density lipoprotein receptor gene.²⁵ A monogenic component is also the main cause of certain heart diseases, some of which are congenital.²⁶ Despite the high penetrance of monogenic diseases, their frequency in the population is very low. Typically, the main cardiovascular diseases and their intermediate phenotypes respond to a polygenic inheritance pattern involving various types of polymorphisms, with each having a small effect that slightly increases the risk. A glossary of key terms related to genetics and epigenetics is presented in the Table. Understanding of the significance of these terms is required to better comprehend the concepts discussed in this article.

Measurement of DNA Sequence Variations and the Methodology for the Analysis of Their Association With Cardiovascular Phenotypes

In recent years, spectacular advances have been made in the technology used to determine DNA variations. Highly accurate, fast, and inexpensive data can now be obtained on the presence of specific genetic variants in an individual's genome. The economic cost and time required depend on the number of genetic variations to be analyzed. The first step in the genome analysis process involves DNA extraction. To analyze genomic DNA, any biological

Table
Glossary of Terms in Genetics and Epigenetics

Term	Description
Allele	Each one of the versions of a polymorphism or gene. An individual typically inherits 2 alleles for each polymorphism, 1 from each parent. Individuals with 2 equal alleles are homozygous for that polymorphism. They are heterozygous if the alleles are different.
Autosome	One of the nonsex chromosomes
Chromatin	Material formed by nucleic acids and proteins seen in the cell nucleus during interphase
Chromosome	Organized package of DNA found in the cell nucleus. Each organism has a different number of chromosomes. Humans have 23 pairs of chromosomes
Codon	A sequence of 3 DNA or RNA bases that specifies a single amino acid in the translation process
DNA	The chemical name of the molecule containing the genetic instructions. It consists of 2 strands that are wound around each other to form a double helix. Bound to each sugar is 1 of 4 bases: adenine (A), cytosine (C), guanine (G), and thymine (T). The 2 strands are bound to each other via well-defined links between specific bases
Epigenetic	Mechanism of the regulation (transcription and translation) of gene expression that does not rely on changes in DNA bases but that operates at a higher level
EWA	Epigenome-wide association. It refers to studies analyzing the methylation of the whole epigenome
Exon	Coding region of a gene
Gene	Unit of inheritance that occupies a specific position in the genome (locus) and has a particular structure
Genome	Basic set of chromosomes containing all genetic information of an individual
Genotype	Sum of an individual's alleles for a specific position
GRS	Genetic risk score. This score can be weighted or unweighted depending on whether the effect size of each polymorphism is evaluated or only the number of risk alleles
GWAS	Genome-wide association study
Histone	Small basic pH proteins, rich in lysine and arginine, that are bound to DNA at the chromatin
Locus	Place occupied by a gene in the genome
Methylation	Addition of methyl groups ($-\text{CH}_3$) to DNA in the form of methylated bases
MicroRNA	Tiny fragment of noncoding RNA with an important regulatory function
Mutation	Any base change introduced in the DNA sequence. Sometimes specifically used to indicate that the allelic frequency of a genetic variation is very low ($< 1\%$)
Nucleotide	Molecule formed by a nitrogenous base, a pentose sugar, and a phosphoric acid group. Basic unit of nucleic acid
Phenotype	Characteristic or observable trait of individuals, resulting from an interaction between their genotype and the environment in which it is expressed. Final cardiovascular disease phenotypes (eg, infarction, stroke) are distinguished from intermediate phenotypes (eg, hypertension, dyslipidemias)
RNA	Molecule formed by a polyribonucleotide of varying lengths that contains uracil instead of thymine. There are 3 types: messenger RNA (mRNA), ribosomal RNA (rRNA), and transfer RNA (tRNA)
Telomere	Extreme end of chromosomes. In humans, the DNA of telomeres is composed of tandem repeats of the sequence TTAGGG

sample containing nucleated cells can be used. Peripheral blood leukocytes are the standard specimen. The DNA extracted from this source is of a sufficient concentration and quality for most subsequent genetic analyses. As a noninvasive alternative for biological sample collection, saliva can be used, but the concentration and quality of the DNA extracted from this specimen might be insufficient for massive genetic analyses, which is why researchers must weigh up the pros and cons of each option.

If the goal is simply to analyze a single polymorphism or a small number in candidate genes, the SNP determination technique used is typically based on fluorescent probes; this technique replaces an earlier one involving restriction enzymes and agarose gels, known as restriction fragment length polymorphism analysis. For polymorphism nomenclature, a system has been adopted to unequivocally identify each genetic variant. In this system, polymorphisms are assigned a number with the prefix “rs” (“reference SNP”).¹⁴ This is an opportune moment to clarify the conceptual difference between mutations and polymorphisms. The difference is academic because the terms are frequently used synonymously. Both refer to DNA base changes, and the difference lies in their frequency. The term mutation refers to uncommon variants (with an allelic frequency < 1%), whereas polymorphism refers to variations with allelic frequencies > 1%. An example of a DNA variant consisting of an A>G change (rs9999) is shown in Figure 1. This change causes 2 alleles (allele A and allele G). These alleles give rise to 3 corresponding genotypes and distinct inheritance models (dominant, recessive, codominant, and additive) can be used when studying their association with disease phenotypes. The figure also shows how to calculate allelic and genotype frequencies and illustrates the Hardy-Weinberg equilibrium test.

To study a greater number of polymorphisms, distinct types of chips are used allowing higher-density genotyping. These chips include those termed “whole-genome” chips, thus named because they include polymorphisms distributed throughout all chromosomes. The first whole-genome genotyping chips contained 10 000 SNPs and were also called 10K chips (with K indicating thousands). Chip density has since increased. Thus, for example, the first

genome-wide association studies (GWASs) performed in the Framingham study, published in 2007, used 100K genotyping chips.²⁷ The density of these chips has subsequently increased even further, and current chips have genotyping densities of more than 1 million SNPs per individual. Statistical analyses of the association between whole-genome genotyping and the cardiovascular phenotype of interest²⁸ frequently use graphs called Manhattan plots, due to their similarity to the skyscrapers of Manhattan Island in New York. Figure 2 shows a Manhattan plot corresponding to a dense GWAS using a 1000K array and an intermediate cardiovascular phenotype. The vertical axis is the $-\log_{10}$ value of the P for association between each SNP and the phenotype of interest, whereas the horizontal axis is the position occupied by each SNP in the chromosome. Each point on the graph is an SNP, so that only those SNPs with very low P values can be seen; the other SNPs are visible as a denser mass in the lower part of the graph. The higher the position of the SNP, the stronger its association with the cardiovascular phenotype (ie, with a lower P value for association). For an association to be considered statistically significant, a nominal P value less than .05 is not used. Instead, this figure is corrected by the number of comparisons made to minimize the occurrence of false positives. The commonly accepted threshold value for a statistically significant association in GWASs is $P < 5 \times 10^{-8}$; this P value is equivalent to a $-\log$ value of 7.25. Some authors also use $P < 10^{-8}$, which is equivalent to a $-\log$ value of 8. When an exploratory GWAS is performed, whose results are to be confirmed in another population, the P values determined are used as suggestive association results for subsequent confirmation. This value is $P < 10^{-5}$. Hundreds of GWASs have been performed with intermediate and final cardiovascular disease phenotypes and the main genes and their corresponding SNPs have been identified for each disease. Given that a detailed list of these polymorphisms is beyond the scope of this review, we recommend that the original publications be consulted or the relevant synthetic reviews.^{6,28}

The GWASs enable us to individually determine the main SNPs associated with the phenotype of interest. To determine their joint contribution, genetic risk scores (GRSs) are used. Figure 3 shows

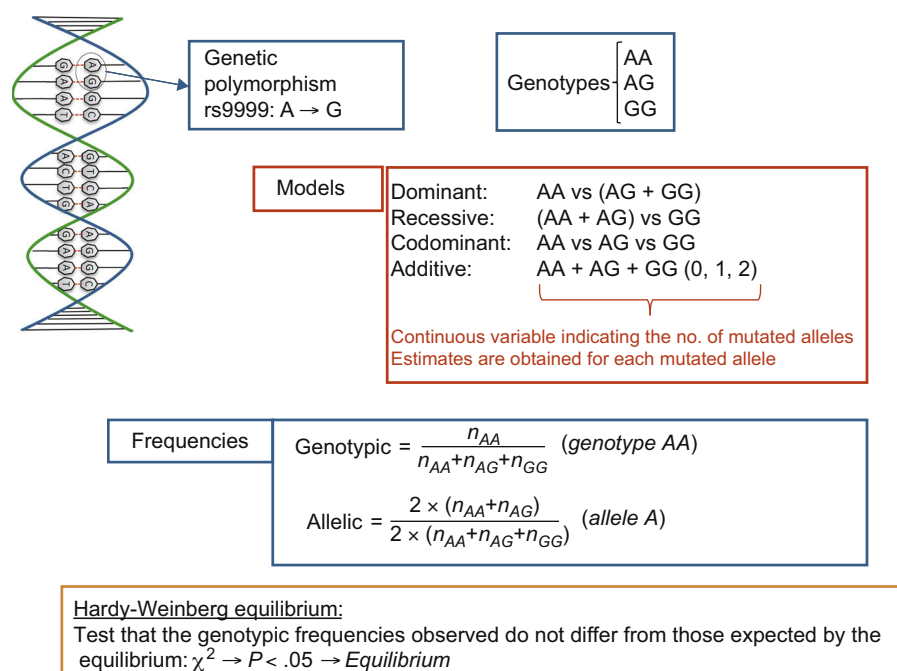


Figure 1. Genetic polymorphism, genotypes, inheritance models, calculation of genotype frequencies, allelic frequencies, and the Hardy-Weinberg principle.

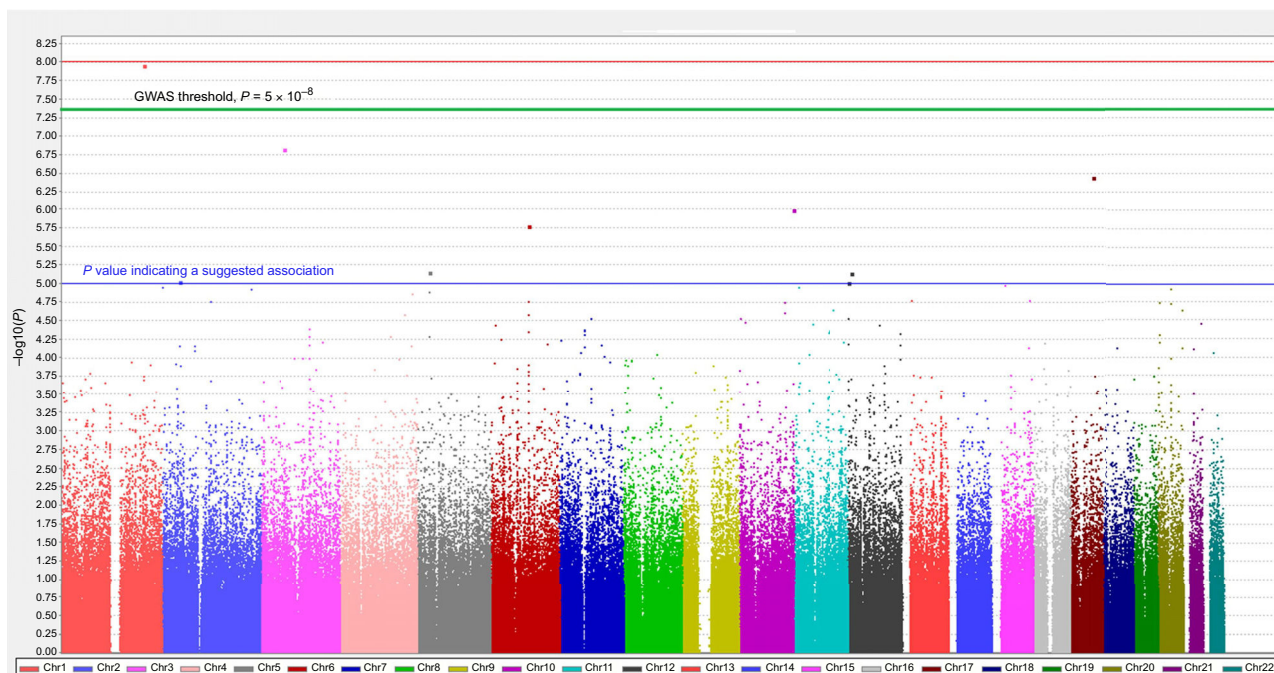


Figure 2. Manhattan plot of a genome-wide association study. Chr, chromosome; GWAS, genome-wide association study.

the calculation of the 2 main types of GRSs: a) unweighted and b) weighted. Numerous studies have analyzed and quantified the influence of various GRSs associated with distinct cardiovascular phenotypes.²⁹ Greater detail on the calculation of GRSs and their advantages and disadvantages can be found in other reviews.³⁰

At the same time, there has also been a considerable leap forward in direct sequencing technology, now known as next-generation sequencing (NGS). This approach, instead of being based on techniques applied in the Human Genome Project (the traditional Sanger capillary electrophoresis sequencing method, considered the first generation), NGS techniques obtain more data at lower cost. The NGS has 3 major improvements³¹: a) it does not require a bacterial cloning step and sequencing libraries are prepared in a cell-free system; b) millions of sequencing reactions are processed in parallel and at the same time; c) base detection is performed cyclically and in parallel. All of these factors greatly improve reliability and reduce cost. Various projects are underway that aim to sequence and map the genomes of thousands of individuals, with the help of a technology called HiSeq X, a high-throughput sequencer from Illumina. The HiSeq X produces 1.8 Tb of sequence data per run in 3 days and is specifically designed for whole-genome sequencing, which requires very high-throughput and simultaneous multiparallel sequencing.³¹ The target of whole-genome sequencing for \$1000 is in sight, and the cost might be even lower with third-generation technologies.³¹ Currently, exome sequencing is already quite advanced in cardiovascular genetic studies³² and will gradually be followed by sequencing of the whole genome.

EPIGENOMICS

Despite the considerable advances in genome sequencing, understanding of the genetic variants in DNA is insufficient to predict disease risk because there are other more dynamic regulatory elements, called the epigenome, that are also able to regulate the expression of DNA sequences. Accordingly, it is essential to study the genome and epigenome jointly to better

understand the molecular bases of the cardiovascular risk phenotypes. The existence of epigenetic regulation helps to explain why the fundamental dogma of biology is not followed (Figure 4). In the classic process shown in panel A, DNA is transcribed to RNA and RNA is translated into protein, whereas panel B shows the more complex currently understood process, with a single DNA able to give rise to different proteins via the influence of distinct epigenetic regulatory processes that go beyond DNA mutations.¹³ Epigenomics is the study of the key functional elements regulating cellular gene expression. In contrast to the genome, which is the same in all somatic cells, the epigenome is specific for each cell type, increasing the complexity of sequencing studies and placing greater emphasis on the origin of the sample used for the analysis. There will be differences between epigenome analysis results if leukocytes or adipose tissue, for example, are analyzed, even when obtained from the same person.²⁰ To analyze the epigenome of different cell types, various consortiums, brought together under the umbrella of the International Human Epigenome Consortium, have taken on the challenge of decoding the epigenome sequences of hundreds of healthy and unhealthy people, specific for each cell type, and are disseminating the results in various publications.³³ Of the numerous types of epigenetic modifications, the most widely studied are methylations and regulations by noncoding RNA such as microRNA and other types of RNA. Histone modifications, another type of epigenetic regulation, are more complex and are still little studied in human epidemiological studies. The huge interest in improving our understanding of epigenetic regulators lies in their dynamicity. In contrast to SNPs in DNA, epigenetic marks can be modified and understanding of the factors involved in this modification might be crucial for the prevention and/or treatment of cardiovascular diseases.

Regulation by Methylation

This type of epigenetic modification of DNA involves the enzymatic addition of a methyl group to the C-5 carbon of cytosine (Figure 5). Most 5-methylcytosines are present on

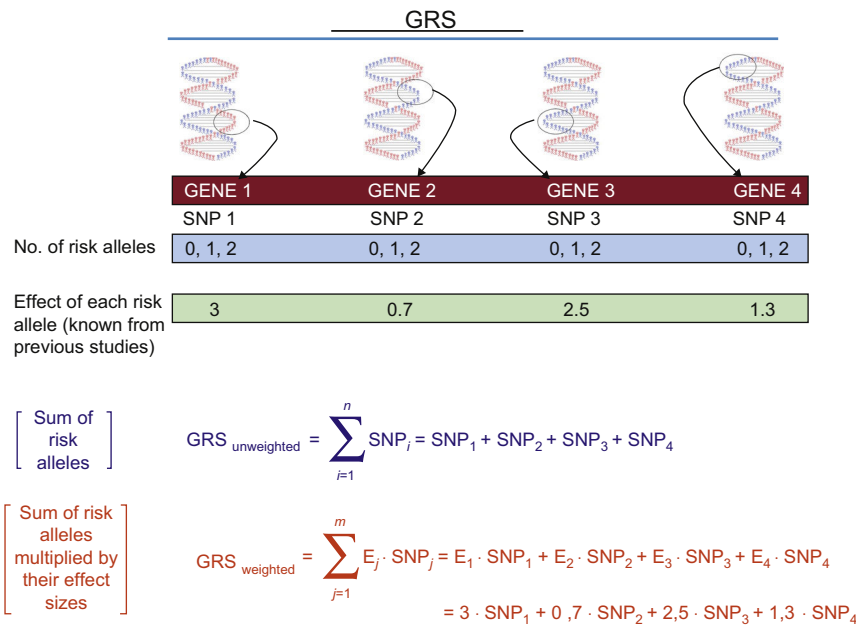


Figure 3. Calculation of unweighted and weighted GRSs. GRS, genetic risk score; SNP, single-nucleotide polymorphism.

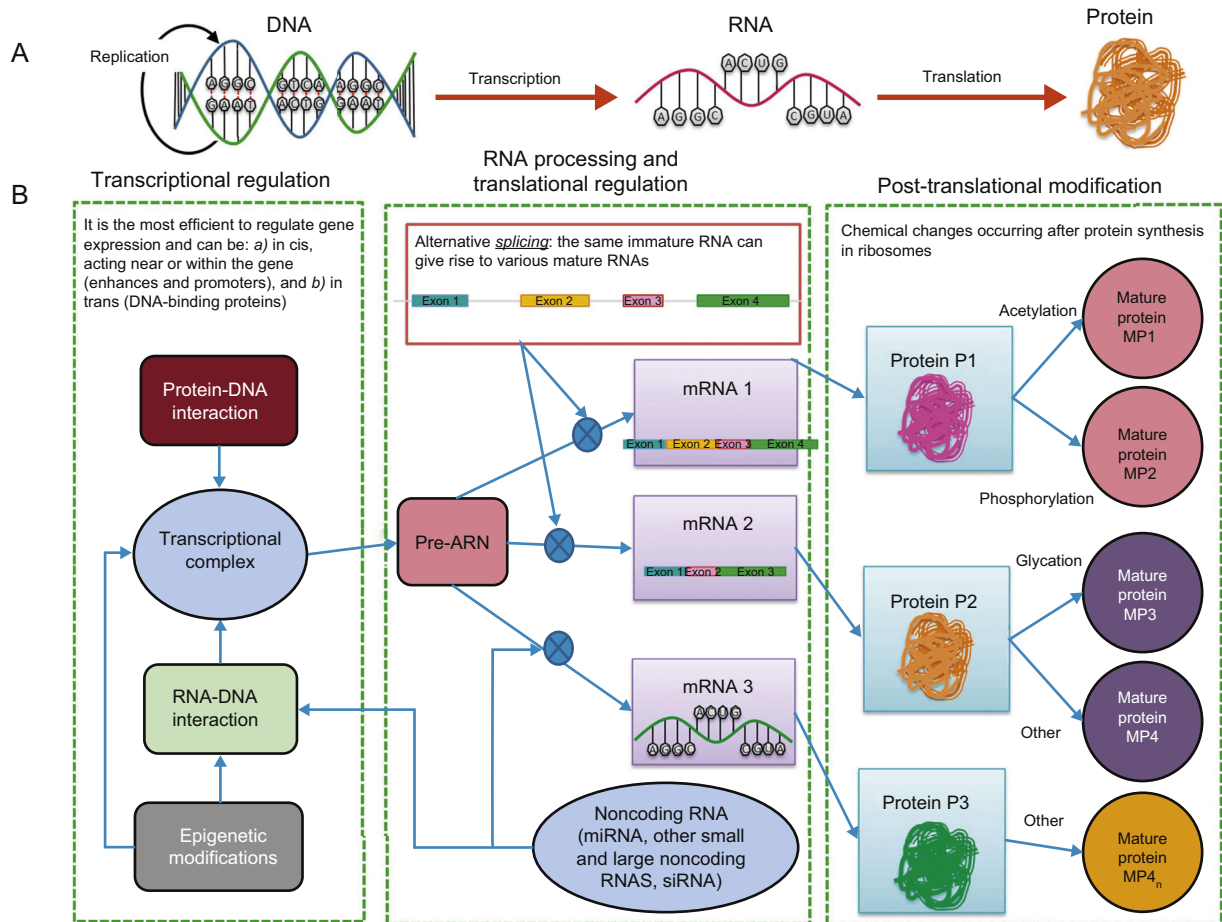


Figure 4. Transcription and translation. A: According to the central dogma of biology, with a single gene giving rise to a single protein. B: The modern view, which includes various additional regulatory elements that allow the same gene to give rise to various proteins. miRNA, microRNA; MP, mature protein; mRNA, messenger RNA; Pre-RNA, precursor RNA; siRNA, silencing RNA.

cytosine-phosphate-guanine (CpG) dinucleotides. The DNA methylation is performed by DNA methyltransferases. These enzymes are classified into 2 groups: maintenance DNA methyltransferases (DNMT1), which maintain methylation patterns, and de novo DNA methyltransferases (DNMT3A and DNMT3B), which perform new methylations. Demethylases are in charge of the inverse process (ie, methyl group elimination) but these mechanisms are less well understood. Greater detail on all of these methylation and demethylation processes can be found in specific themed reviews.^{20,34,35} The CpG dinucleotides are not uniformly distributed throughout the genome but are concentrated in certain regions. Areas enriched in these dinucleotides (> 60%) are called CpG islands. These islands are usually concentrated between the promoter and the transcription start site. Methylation-dependent epigenetic regulation is complex but methylation of regulatory elements in genes, such as promoters, enhancers, isolators, and repressors, generally suppresses gene function (Figure 5).³⁵ There are various ways to analyze the methylation of a specific gene or the genome as a whole. Accordingly, as discussed for whole-genome genotyping, in epigenome-wide association (EWA) studies of methylation, specific chips are used to analyze the most important regions of the genome. The initial 450K methylation chip of Illumina has recently been replaced with a chip with greater coverage, 850K, also produced by Illumina.³⁶ To analyze associations, Manhattan plots are also used but, in these studies, the analyses evaluate the *P* values for the association between the methylation and the specific phenotype.³⁷ Although various EWA studies have been published, the agreement among the results of these studies is still low. Considering other types of methylation, a review has analyzed the results of 31 articles focused on methylation and plasma lipids that included 8027 participants.³⁸ Generally, there were no firm associations between general DNA methylation and plasma lipids. Regarding the methylation of specific genes, consistent results were found with *ABCG1*, *CPT1A*, *TNNT1*, *MIR33B*, *SREBF1*, and *TNIP*,³⁸ which is why more in-depth study is required of these relationships and their modulators. Because dietary factors and smoking strongly influence DNA methylation,³⁹ their effects need to be strongly considered in EWA studies.

Interest has also recently grown into hydroxymethylation regulation of DNA.⁴⁰ This process is carried out by TET (ten-eleven translocation methylcytosine dioxygenase) enzymes, which recognize certain previously methylated cytosines and oxidize them, converting 5-methylcytosine to 5-hydroxymethylcytosine.

Hydroxymethylated genes appear to be associated with increased transcription. Because this is a relatively new field, many more studies are required to determine the influence of hydroxymethylation on cardiovascular risk.

Regulation by Methylcytosine-phosphate-guanine-binding Proteins

This mechanism involves proteins or protein complexes that specifically bind to methylated CpG regions and indirectly block the binding of transcription factors by limiting their access to regulatory elements.³⁵ These proteins contain conserved methylated DNA-binding domains called methyl-binding domains (MBDs). The first protein identified was MeCP2, which was followed by other proteins, such as MBD1, MBD2, and MBD3, involved in transcriptional repression, the same as MeCP2. Other MBD proteins have distinct functions. This type of regulation is generally complex.³⁵

Regulation by Noncoding RNA. MicroRNA and Others

A large number of noncoding RNA transcripts have been identified. Their function was unknown for many years but progress is now being made in understanding their roles.⁴¹ These RNA do not encode proteins but play important regulatory roles in many processes. They are classified as short noncoding RNA (< 200 base pairs) and long noncoding RNA (> 200 base pairs).¹⁹ The shortest and best studied are microRNA (~20–25 base pairs). microRNA are encoded in the genome, either in the introns within another coding gene or in the intergenic spaces. They are transcribed as a larger double-stranded primary transcript (pri-miR) by RNA polymerase II. Subsequently, the nuclear enzymes Drosha (also known as ribonuclease III) and Pasha convert this precursor into a double-stranded 70-nucleotide microRNA precursor (pre-miR), and a mechanism involving the protein exportin transports it to the cytoplasm. Finally, the enzyme Dicer processes this precursor into a double-stranded 22-nucleotide microRNA. This duplex is unrolled and 1 of the 2 strands is incorporated into the RISC (RNA-induced silencing complex), which comprises the protein argonaute, among others.⁴² The microRNA species incorporated in the RISC are able to bind to the 3'-UTR of target messenger RNA and block translation. The other microRNA strand is degraded. Figure 6 shows a schematic

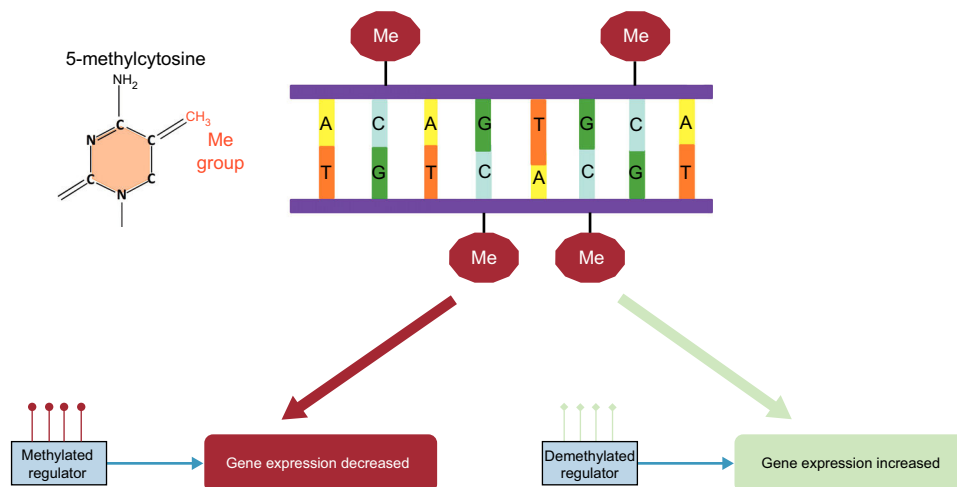


Figure 5. DNA methylation and its effect on gene expression. Me, methyl.

representation of the binding of microRNA to its target messenger RNA that blocks protein production and a comparison of the process without microRNA interference. Multiple microRNAs are involved in the regulation of the distinct cardiovascular phenotypes. These microRNAs can be analyzed in specific tissues and their expression related to certain phenotypic characteristics of tissues, such as, for example, the relationship between the expression of the microRNAs miR-1, miR-133, and miR-208 and the development of ventricular hypertrophy and heart failure. Additionally, microRNA can be found in the plasma circulation and analysis of their concentrations can indicate various diseases and processes. For example, the plasma levels of miR-1, miR-133a, miR-133b, and miR-499-5p are elevated after an acute myocardial infarction.⁴² This emerging research field, despite its considerable spread, requires better standardization of techniques and processes because it is subject to highly dynamic variation. After this standardization, multiple applications will no doubt be found for this fascinating world of epigenomic regulation.

Regulation by Histone Modification

This type of epigenetic regulation is more complex and less well-analyzed in human epidemiological studies. The main post-translational modifications that can take place in histones include acetylation, phosphorylation, methylation, and ubiquitination and influence the compaction of chromatin.²⁰ Histones have a globular carboxy terminus and an amino terminal tail, which is where the modifications occur. However, these modifications do not occur in any site but respond to well-defined regulatory codes. Thus, for example, methylation occurs in lysine (abbreviated as K) and arginine (R) residues, acetylation in lysine residues, ubiquitination in lysine, and phosphorylation in serines (S) and threonines (T). The outcome is also affected by the type of histone being modified.³⁵ This selective modification of histones is illustrated in Figure 7. Despite the difficulties, several studies have reported that HDAC3 (histone deacetylase 3) plays a critical role in endothelial function, whereas HDAC7 (histone deacetylase

7) is important for the function of vascular smooth muscle cells.³⁵ In addition, histone enzymes tend to be more underexpressed than overexpressed in metabolic diseases.⁴³ However, standardization is still required of the differences in their expression according to tissue and other complex regulators that also participate in the processes.³⁵

GENOMIC AND EPIGENOMIC INTEGRATION. GENE-ENVIRONMENT INTERACTIONS

Although the genomic and epigenomic concepts have been presented separately, the reality is that they considerably interact. For example, methylation depends on certain genetic polymorphisms that can be found in the sequence.⁴⁴ Similarly, the binding of a microRNA to a target messenger RNA will differ if there is a specific base sequence or a genetic polymorphism.⁴⁴ Thus, some of our interesting results can serve as an example. In our work, we analyzed a polymorphism at the binding site of a microRNA (in this case, miR410) in the lipoprotein lipase gene on plasma triglyceride concentrations and its subsequent modulation by the Mediterranean diet to determine stroke risk.⁴⁵ All of these factors indicate the need for integrated studies, not only of -omics, but also of environmental factors such as diet and other life style variables, to better understand the molecular bases of cardiovascular risk and their regulation. To study gene-environment interactions, the proposed approach has been to explore the statistical interaction, instead of focusing on biological interactions (see a review⁴⁶ to better understand the differences between them). However, the publication of an important biological, but not statistically significant, interaction between a GRS of cardiovascular risk and life style that determines the incidence of cardiovascular disease⁴⁷ has helped to support the concept of a gene-environment interaction, understood as the existence of an environmental factor that can modify genetic susceptibility, and has led to new studies being planned from this perspective.

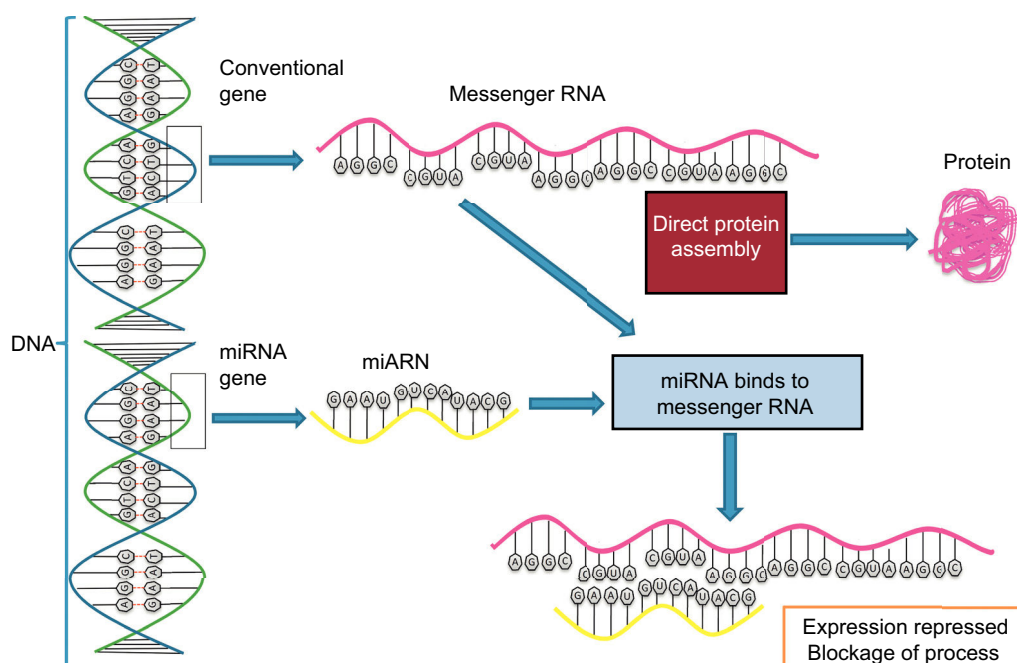


Figure 6. Binding of a microRNA to messenger RNA and blockage of translation. miRNA, microRNA.

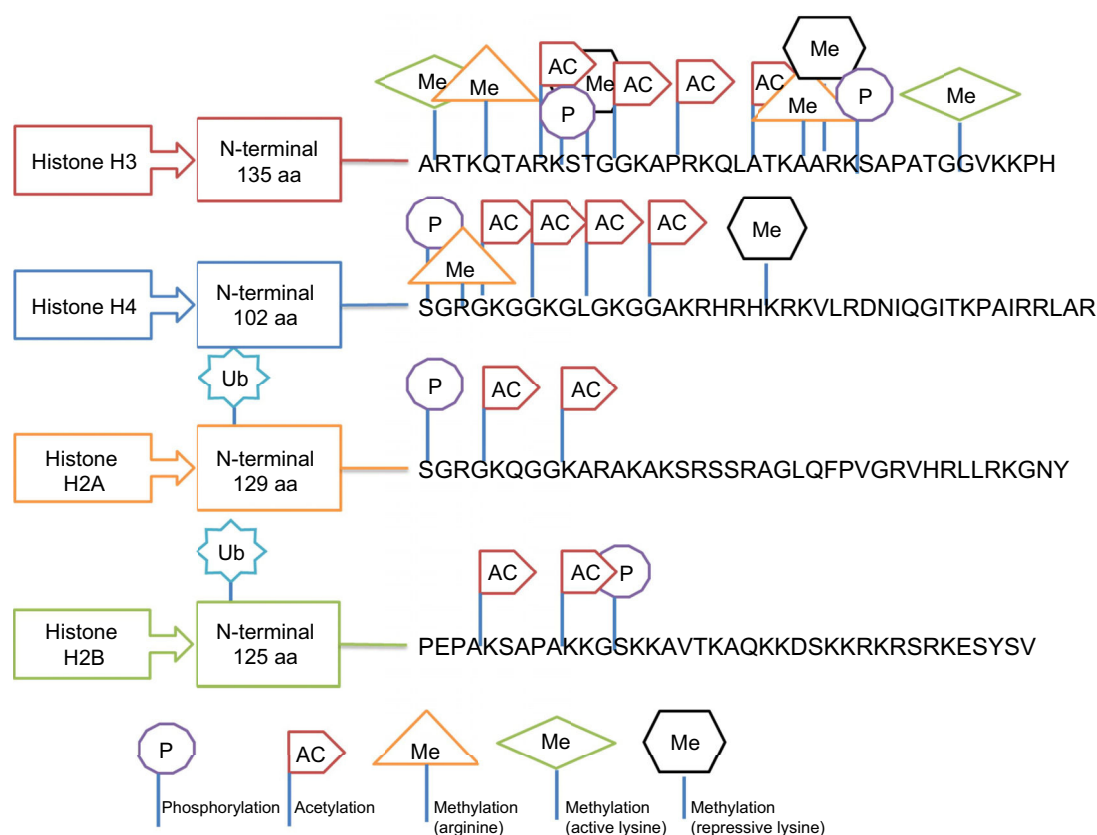


Figure 7. Epigenetic modifications of histones (amino acids are represented by the code of the corresponding letter).

FUNDING

This work has been supported by the Spanish Ministry of Health (*Instituto de Salud Carlos III*) and by the Ministry of Economy, Industry and Competitiveness (CIBER 06/03, CNIC-06/2007, RTIC G03/140, SAF2016-80532-R, AES_FIS_2016), and the TV3 Marathon Foundation (538/U/2016).

CONFLICTS OF INTEREST

None declared.

REFERENCES

- Chaisson MJ, Wilson RK, Eichler EE. Genetic variation and the de novo assembly of human genomes. *Nat Rev Genet.* 2015;16:627–640.
- MacRae CA, Vasan RS. The future of genetics and genomics: closing the phenotype gap in precision medicine. *Circulation.* 2016;133:2634–2639.
- Vasan RS, Benjamin EJ. The future of cardiovascular epidemiology. *Circulation.* 2016;133:2626–2633.
- Price MJ. Genetic considerations. *Adv Cardiol.* 2012;47:100–113.
- Pereira NL, Sargent DJ, Farkouh ME, Rihal CS. Genotype-based clinical trials in cardiovascular disease. *Nat Rev Cardiol.* 2015;12:475–487.
- Kathiresan S, Srivastava D. Genetics of human cardiovascular disease. *Cell.* 2012;148:1242–1257.
- Musunuru K, Kathiresan S. Surprises from genetic analyses of lipid risk factors for atherosclerosis. *Circ Res.* 2016;118:579–585.
- Collins FS, Varmus H. A new initiative on precision medicine. *N Engl J Med.* 2015;372:793–795.
- Houser SR. The American Heart Association's New Institute for Precision Cardiovascular Medicine. *Circulation.* 2016;134:1913–1914.
- Joyner MJ. Precision medicine, cardiovascular disease and hunting elephants. *Prog Cardiovasc Dis.* 2016;58:651–660.
- Klein G, Klein E. The rise and fall of central dogmas. *Oncimmunology.* 2016;5:e1043071.
- Guttman AE, Collins FS. Genomic medicine—a primer. *N Engl J Med.* 2002;347:1512–1520.
- Feero WG, Guttman AE, Collins FS. Genomic medicine—an updated primer. *N Engl J Med.* 2010;362:2001–2011.
- Attia J, Ioannidis JP, Thakkinstian A, et al. How to use an article about genetic association: A: Background concepts. *JAMA.* 2009;301:74–81. Erratum in: *JAMA.* 2009;301:1024.
- Attia J, Ioannidis JP, Thakkinstian A, et al. How to use an article about genetic association: B: Are the results of the study valid? *JAMA.* 2009;301:191–197.
- Attia J, Ioannidis JP, Thakkinstian A, et al. How to use an article about genetic association: C: What are the results and will they help me in caring for my patients? *JAMA.* 2009;301:304–308.
- Yan H, Tian S, Slager SL, Sun Z, Ordog T. Genome-wide epigenetic studies in human disease: a primer on -omic technologies. *Am J Epidemiol.* 2016;183: 96–109.
- Wolyniak MJ, Bemis LT, Prunuske AJ. Improving medical students' knowledge of genetic disease: a review of current and emerging pedagogical practices. *Adv Med Educ Pract.* 2015;6:597–607.
- Musunuru K, Hickey KT, Al-Khatib SM, et al. Basic concepts and potential applications of genetics and genomics for cardiovascular and stroke clinicians: a scientific statement from the American Heart Association. *Circ Cardiovasc Genet.* 2015;8:216–242.
- Guardiola M, Vallvé JC, Zaina S, Ribalta J. Epigenetics in atherosclerosis. *Clin Invest Arterioscler.* 2016;28:102–119.
- Annunziato A. DNA packaging: nucleosomes and chromatin. *Nature Education.* 2008;1:26.
- Fischle W, Wang Y, Allis CD. Histone and chromatin cross-talk. *Curr Opin Cell Biol.* 2003;15:172–183.
- Gilson E, Ségal-Bendirdjian E. The telomere story or the triumph of an open-minded research. *Biochimie.* 2010;92:321–326.
- Yeh JK, Wang CY. Telomeres and telomerase in cardiovascular diseases. *Genes (Basel).* 2016. <http://dx.doi.org/10.3390/genes7090058>.
- Ramasamy I. Update on the molecular biology of dyslipidemias. *Clin Chim Acta.* 2016;454:143–185.
- Marian AJ, Van Rooij E, Roberts R. Genetics and genomics of single-gene cardiovascular diseases: common hereditary cardiomyopathies as prototypes of single-gene disorders. *J Am Coll Cardiol.* 2016;68:2831–2849.
- Kathiresan S, Manning AK, Demissie S, et al. A genome-wide association study for blood lipid phenotypes in the Framingham Heart Study. *BMC Med Genet.* 2007;8(Suppl 1):S1–S17.
- Elosua R, Lluís C, Lucas G. Estudio del componente genético de la cardiopatía isquémica: de los estudios de ligamiento al genotipado integral del genoma. *Rev Esp Cardiol Supl.* 2009;9(B):24–38.
- Assimes TL, Salfati EL, Del Gobbo LC. Leveraging information from genetic risk scores of coronary atherosclerosis. *Curr Opin Lipidol.* 2017;28:104–112.

30. Malovini A, Bellazzi R, Napolitano C, Guffanti G. Multivariate methods for genetic variants selection and risk prediction in cardiovascular diseases. *Front Cardiovasc Med*. 2016;3:17.
31. Park ST, Kim J. Trends in next-generation sequencing and a new era for whole genome sequencing. *Int Neurourol J*. 2016;20(Suppl 2):S76–S83.
32. Brown TL, Meloche TM. Exome sequencing—a review of new strategies for rare genomic disease research. *Genomics*. 2016;108:109–114.
33. Kundaje A, Meuleman W, Ernst J, et al. Roadmap Epigenomics Consortium. Integrative analysis of 111 reference human epigenomes. *Nature*. 2015;518:317–330.
34. Khyzha N, Alizada A, Wilson MD, Fish JE. *Epigenetics of atherosclerosis: emerging mechanisms and methods*. Trends Mol Med. 2017; 23:332–347.
35. Mazzio EA, Soliman KF. Basic concepts of epigenetics: impact of environmental signals on gene expression. *Epigenetics*. 2012;7:119–130.
36. 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;8:389–399.
37. Rask-Andersen M, Martinsson D, Ahsan M, et al. Epigenome-wide association study reveals differential DNA methylation in individuals with a history of myocardial infarction. *Hum Mol Genet*. 2016;25:4739–4748.
38. Braun KV, Voortman T, Dhana K, et al. The role of DNA methylation in dyslipidaemia: A systematic review. *Prog Lipid Res*. 2016;64:178–191.
39. Piyasena C, Cartier J, Provenal N, et al. Dynamic changes in DNA methylation occur during the first year of life in preterm infants. *Front Endocrinol (Lausanne)*. 2016;7:158.
40. Pauwels S, Duca RC, Devlieger R, et al. Maternal methyl-group donor intake and global DNA (hydroxy)methylation before and during pregnancy. *Nutrients*. 2016. <http://dx.doi.org/10.3390/nu8080474>.
41. Viereck J, Thum T. Circulating noncoding RNAs as biomarkers of cardiovascular disease and injury. *Circ Res*. 2017;120:381–399.
42. Paul P, Chakraborty A, Sarkar D, et al. Interplay between miRNAs and human diseases: a review. *J Cell Physiol*. 2017. <http://dx.doi.org/10.1002/jcp.25854>.
43. Shao Y, Chernaya V, Johnson C, et al. Metabolic diseases downregulate the majority of histone modification enzymes, making a few upregulated enzymes novel therapeutic targets—“sand out and gold stays”. *J Cardiovasc Transl Res*. 2016;9:49–66.
44. Ma Y, Ordovas JM. The integration of epigenetics and genetics in nutrition research for CVD risk factors. *Proc Nutr Soc*. 2016;1–14.
45. Corella D, Sorlí JV, Estruch R, et al. MicroRNA-410 regulated lipoprotein lipase variant rs13702 is associated with stroke incidence and modulated by diet in the randomized controlled PREDIMED trial. *Am J Clin Nutr*. 2014;100:719–731.
46. Corella D, Ordovas JM. Nutrigenomics in cardiovascular medicine. *Circ Cardiovasc Genet*. 2009;2:637–651.
47. Khera AV, Emdin CA, Drake I, et al. Genetic risk, adherence to a healthy lifestyle, and coronary disease. *N Engl J Med*. 2016;375:2349–2358.