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]

[a] Departamento de Medicina Preventiva y Salud Pública, Facultad de Medicina, Universidad de Valencia, Valencia, Spain
[b] CIBER Fisiopatología de la Obesidad y Nutrición, Instituto de Salud Carlos III, Madrid, Spain
[c] Departamento de Epidemiología Cardiovascular y Genética de Poblaciones, Centro Nacional de Investigaciones Cardiovasculares (CNIC), Madrid, Spain
[d] Instituto Madrileño de Estudios Avanzados (IMDEA) Alimentación, Madrid, Spain
[e] Nutrition and Genomics Laboratory, Jean Mayer USDA Human Nutrition Research Center on Aging, Tufts University, Boston, Massachusetts, United States

A B S T R A C T

Keywords:
Genetics
Epigenetics
Polymorphism
Methylation

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

R E S U M E N

Palabras clave:
Genética
Epigenética
Polimorfismo
Metilación

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.

* 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).

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