JOURNAL OF THE AMERICAN COLLEGE OF CARDIOLOGY
© 2016 BY THE AMERICAN COLLEGE OF CARDIOLOGY FOUNDATION
PUBLISHED BY ELSEVIER

VOL. 68, NO. 25, 2016
ISSN 0735-1097/$36.00
http://dx.doi.org/10.1016/j.jacc.2016.09.968

FOCUS SEMINAR: GENETICS

STATE-OF-THE-ART REVIEW

Genetics and Genomics of Single-Gene Cardiovascular Diseases

Common Hereditary Cardiomyopathies as Prototypes of Single-Gene Disorders

Ali J. Marian, MD,[a] Eva van Rooij, PhD,[b,c] Robert Roberts, MD[d]

ABSTRACT

This is the first of 2 review papers on genetics and genomics appearing as part of the series on "omics." Genomics pertains to all components of an organism's genes, whereas genetics involves analysis of a specific gene or genes in the context of heredity. The paper provides introductory comments, describes the basis of human genetic diversity, and addresses the phenotypic consequences of genetic variants. Rare variants with large effect sizes are responsible for single-gene disorders, whereas complex polygenic diseases are typically due to multiple genetic variants, each exerting a modest effect size. To illustrate the clinical implications of genetic variants with large effect sizes, 3 common forms of hereditary cardiomyopathies are discussed as prototypic examples of single-gene disorders, including their genetics, clinical manifestations, pathogenesis, and treatment. The genetic basis of complex traits is discussed in a separate paper. (J Am Coll Cardiol 2016;68:2831–49) © 2016 by the American College of Cardiology Foundation.

The human nuclear genome is composed of 3.2 billion base pairs and 20,576 protein-coding genes, which are arranged in 22 pairs of somatic and a pair of sex (X and Y) chromosomes (NCBI Homo sapiens Annotation Release 107). Chromosome 1 is the largest chromosome, containing approximately ~249 million base pairs and about 4,000 genes (Assembly hg38, UCSC Genome Browser, University of California, Santa Cruz, California). Chromosome 21 is the smallest, with about 48 million base pairs and 250 protein-coding genes. Over 90% of the genome is transcribed, predominantly into noncoding ribonucleic acids (ncRNAs), and only ~1% of the genome is translated into protein.

The nuclear genome also contains about 18,000 genes that are transcribed into ncRNAs. The ncRNAs are transcribed from active chromatin, polyadenylated, and capped, but typically are not translated into proteins. The ncRNAs are commonly classified into small ncRNAs, which are usually <200 nucleotides in length, and long noncoding ribonucleic acids (lncRNAs), which are >200 nucleotides long. Microribonucleic acids (miRNAs) are the best-studied ncRNAs. They are initially transcribed as longer transcripts, and then are cleaved into mature 22-nucleotide-long miRNAs. miRNAs repress gene expression by binding to a recognition sequence, typically within the 3′ untranslated regions of target messenger ribonucleic acids (mRNAs). lncRNAs regulate gene expression through a broad range of effects, including forming complexes with proteins and as sponges for other transcripts.

The human genome also contains pseudogenes, which no longer code for proteins. In addition, repetitive deoxyribonucleic acid (DNA) elements occupy more than 50% of the human nuclear genome.

Listen to this manuscript's audio summary by *JACC* Editor-in-Chief Dr. Valentin Fuster.

From the [a]Center for Cardiovascular Genetics, Brown Foundation Institute of Molecular Medicine, The University of Texas Health Science Center, and Texas Heart Institute, Houston, Texas; [b]Hubrecht Institute, KNAW and University Medical Center Utrecht, Utrecht, the Netherlands; [c]Department of Cardiology, University Medical Center Utrecht, Utrecht, the Netherlands; and the [d]University of Arizona College of Medicine, Phoenix, Arizona. Dr. van Rooij has served on the advisory board for miRagen Therapeutics Inc. All other authors have reported that they have no relationships relevant to the contents of this paper to disclose.

Manuscript received August 2, 2016; revised manuscript received September 14, 2016, accepted September 19, 2016.

ABBREVIATIONS AND ACRONYMS

AC = arrhythmogenic cardiomyopathy

ARVC = arrhythmogenic right ventricular cardiomyopathy

DCM = dilated cardiomyopathy

GV = genetic variant

HCM = hypertrophic cardiomyopathy

lncRNA = long noncoding ribonucleic acid

LOF = loss-of-function

miRNA = microribonucleic acid

SNV = single-nucleotide variant

Functions of the repetitive elements and noncoding regions of the nuclear genome are largely unknown.

DNA is wrapped around a set of octomeric protein complexes comprising 4 highly conserved core (H2A, H2B, H3, and H4) histones, and 2 linker (H1 and H5) histones (**Central Illustration**). This compacted DNA and protein complex is referred to as chromatin. Histones pack the DNA into units of approximately 150 base pairs, referred to as nucleosomes. Tightly packed nucleosomes are referred to as heterochromatin, and are inaccessible to the transcription machinery (inactive transcription). In contrast, loosely packed nucleosomes, referred to as euchromatin, are accessible to the transcription machinery, and hence are actively transcribed.

Histones undergo extensive modifications, including acetylation and methylation, which regulate the chromatin open and closed states, and hence access of the transcription machinery to DNA. For example, trimethylation of lysine residue 27 on histone H3 (H3K27me3) is a chromatin marker for suppression of gene expression. In contrast, acetylation of the same residue (H3K27ac) marks the chromatin for active transcription. Given the large number of residues that could undergo post-translational modifications and various forms of modification, histones are considered major regulators of gene expression. Post-translational histone changes and chemical modifications to DNA (not nucleotide changes), such as CpG methylation, are collectively considered the epigenome.

STRUCTURE OF A GENE

About 1% of the genome, containing approximately 30 million base pairs, codes for proteins. Each protein-coding gene has a 5′ transcriptional regulatory region; protein-coding segments referred to as exons; intervening regions between exons, called introns; and a 3′ untranslated region or regulatory region (**Central Illustration**). Proteins that bind to enhancers, silencers, and promoters proximal to the transcription initiation site regulate transcription. The primary transcripts of genes are spliced to exclude introns and produce mRNAs. Splicing of each primary transcript is not uniform, and often multiple splice variants are generated, of which 1 is the predominant isoform. Each unit of 3 bases, referred to as a *codon,* encodes a specific amino acid. There are 61 amino acid codons and 3 stop codons in the nuclear genome. Thus, each amino acid has multiple codons.

MITOCHONDRIAL GENOME

Each cell also contains mitochondrial deoxyribonucleic acid (mtDNA), a circular DNA composed of 16,700 nucleotides. The mitochondrial genome contains 37 genes, which code for 13 proteins, 2 ribosomal RNAs, and 22 transfer RNAs. Each cell contains a large number of mitochondria, and each mitochondrion typically contains several copies of mtDNA. The codons are largely identical between the nuclear and mitochondrial genomes, except that there are 60 codons for amino acids and 4 for stop codons in the mtDNA. In addition, transcription of mtDNA is continuous, as opposed to the discontinuous transcription of the nuclear genome.

GENERATION OF GENETIC VARIANTS

The replication machinery introduces rare errors during each round of DNA replication, which are empirically calculated to occur at a rate of $\sim 1.1 \times 10^{-8}$ per base pair per generation (1-3). Given the size of the human genome, each DNA replication (meiosis) introduces about 40 to 60 new genetic variants (GVs). Accordingly, each newborn adds 40 to 60 new GVs to the human genetic pool as de novo variants (i.e., absent in the parents). The explosive growth of the human population during the last thousand years or so has introduced a massive number of GVs into the population genetic pool, rendering humans exceedingly diverse at the genetic level (**Table 1**). The mutation rate of the mtDNA is several orders of magnitude higher, likely because of higher oxidative stress, and compromised function of the DNA replication and repair system.

Each nuclear genome contains approximately 4 million GVs, of which ~3.5 million are single-nucleotide variants (SNVs), also referred to as single-nucleotide polymorphisms, and several thousand small insertions/deletions, referred to as indels (4-7). In addition, the nuclear genome contains large insertions, deletions, and rearrangements, which are referred to as structural variants. Structural variants that increase or decrease the number of chromosome segments or genes are referred to as copy number variants. The vast majority of variants in each genome are rare. In addition, rare variants are typically population-specific and hence, vary significantly among people with different ethnic backgrounds.

Approximately 12,000 SNVs change the amino acid sequence and are referred to as nonsynonymous SNVs. Computational programs that incorporate population frequencies of the variants and evolutionary conservation of the involved codons, among