Journal of the American College of Cardiology
© 2013 by the American College of Cardiology Foundation
Published by Elsevier Inc.

Vol. 61, No. 20, 2013
ISSN 0735-1097/$36.00
http://dx.doi.org/10.1016/j.jacc.2012.12.054

STATE-OF-THE-ART PAPER

Genomics in Cardiovascular Disease

Robert Roberts, MD,*† A. J. Marian, MD,‡ Sonny Dandona, MD,§ Alexandre F. R. Stewart, PHD‖

Ottawa, Ontario, and Montreal, Quebec, Canada; and Houston, Texas

A paradigm shift toward biology occurred in the 1990s and was subsequently catalyzed by the sequencing of the human genome in 2000. The cost of deoxyribonucleic acid (DNA) sequencing has gone from millions to thousands of dollars with sequencing of one's entire genome costing only $1,000. Rapid DNA sequencing is being embraced for single gene disorders, particularly for sporadic cases and those from small families. Transmission of lethal genes such as associated with Huntington's disease can, through in vitro fertilization, avoid passing it on to one's offspring. DNA sequencing will meet the challenge of elucidating the genetic predisposition for common polygenic diseases, especially in determining the function of the novel common genetic risk variants and identifying the rare variants, which may also partially ascertain the source of the missing heritability. The challenge for DNA sequencing remains great, despite human genome sequences being 99.5% identical, the 3 million single nucleotide polymorphisms responsible for most of the unique features add up to 40 to 60 new mutations per person which, for 7 billion people, is 300 to 400 billion mutations. It is claimed that DNA sequencing has increased 10,000-fold while information storage and retrieval only 16-fold. The physician and health user will be challenged by the convergence of 2 major trends, whole genome sequencing, and the storage/retrieval and integration of the data. (J Am Coll Cardiol 2013;61:2029–37) © 2013 by the American College of Cardiology Foundation

Captain Cook wrote in his log upon reaching Australia that "I have not only travelled farther than any other man, but I have travelled as far as man can travel" (1). Thus, by the 18th century, all the continents had now been discovered and named. It appeared logical and perhaps appropriate for mankind to pursue the inner treasures of the planet. This coincided with the industrial revolution that led to the harnessing of energy from coal, electricity, and oil as well as the discovery of all the marvelous elements including uranium, which enabled many human endeavors, from cancer therapy to the invention of the atomic bomb. While this trend continues, in the 1990s a major worldwide shift occurred in which mankind became interested in the inner workings of human biology. The word "biology" is today often associated with excitement and activity, not just in science but also in medicine and commerce. This revolutionary concept received a major boost with the sequencing of the human genome in 2000 (2). In fact, sequencing of the human genome may be to the 21st century as invention of the vowels and development of democracy was to the 6th century BC or the industrial revolution was to the 18th century.

The Human Genome: New Developments

The double stranded human genome of each cell contains 6.4 billion nucleotides. While proteins are the molecules that do the work, only about 1% of the human genome sequences are designated to encode messenger ribonucleic acids (RNAs)for protein coding (3). Until recently, most of deoxyribonucleic acid (DNA) was considered junk (3), but we now know that virtually all of DNA is transcribed into RNA (3). The ENCODE (Encyclopedia of DNA Elements) project has enabled us to assign biochemical functions for 80% of the genome (4). It is of note that only a small proportion of the transcribed RNAs are translated into protein with the remainder performing a host of functions, affecting those sequences (genes) that encode for protein. These RNAs that do not code for protein are as a group referred to as noncoding RNA. Most genes coding for protein are in some way regulated by these noncoding RNAs (5). These noncoding RNAs are very promiscuous—each RNA can affect multiple different genes on the same or different chromosomes.

The Source of Human Genetic Biodiversity

All genomes from all species share most of their DNA sequences, having acquired them over a 3.8-billion-year evolutionary history since the origin of life. Despite the common sequence ancestry, each individual genome within each species has maintained itself as unique. The develop-

From the *University of Ottawa Heart Institute, Ottawa, Ontario, Canada; †John & Jennifer Ruddy Canadian Cardiovascular Genetics Centre, Ottawa, Ontario, Canada; ‡Institute of Molecular Medicine, Center for Cardiovascular Genetic Research, University of Texas Health Sciences Center, Houston, Texas; §Department of Medicine, McGill University, Montreal, Quebec, Canada; and the ‖John & Jennifer Ruddy Canadian Cardiovascular Genetics Centre, and University of Ottawa Heart Institute, Ottawa, Ontario, Canada. Dr. Roberts receives grant support from CIHR #MOP82810 (RR)/Canada, CIHR #MOP77682 (AFRS)/Canada, and CFI #11966 (RR)/Canada; and is a consultant to Cumberland Pharmaceuticals. Dr. Marian receives grant support from R01-088498/PHS HHS/United States, R21 AG038597-01/AG/NIA NIH HHS/United States, and R34HL105563/HL/NHLBI NIH HHS/United States. Dr. Stewart receives grant support from CIHR #MOP82810 (RR) & CIHR #MOP77682 (AFRS)/Canada. All other authors have reported that they have no relationships relevant to the contents of this paper to disclose.

Manuscript received November 12, 2012; revised manuscript received January 29, 2013, accepted February 19, 2013.

JACC Vol. 61, No. 20, 2013
May 21, 2013:2029–37

Abbreviations and acronyms

CAD = coronary artery disease

DNA = deoxyribonucleic acid

DSV = deoxyribonucleic acid sequence variants

GWAS = genome-wide association studies

NGS = next generation DNA sequencing

nsSNV = nonsynonymous single-nucleotide variations

RNA = ribonucleic acid

SNP = single nucleotide polymorphism

WES = whole exome sequencing

WGS = whole genome sequencing

ment of biodiversity and unique sequences of each genome whether within or between species is due primarily to the errors in the process of copying DNA. Copying errors during the replication of one's DNA induce primarily single base changes through substitution of a single base (nucleotide) for another (e.g., thymine for adenine). These substitutions are passed on from generation to generation and are referred to as single nucleotide polymorphisms (SNPs). These SNP substitutions account for 94% of the errors from copying or replicating DNA, while deletions of 1 to 4 bp account for 4.5%, and the remainder are due to insertions of 1 to 4 bp (6,7). Other types of DNA variation exist such as chromosomal rearrangements, duplications (copy number variants), and translocations. The mutations induced by DNA copying errors, if beneficial, are conserved and their frequency increases, while deleterious mutations remain rare or are eliminated. Fortunately, many of these SNPs have modest to minimal effects or are neutral. The human DNA (6 billion bases) replicates itself every few days, and although it only makes 1 error per 1 billion bases created, it can accumulate a significant number of mutations over generations. Kruglyayk and Nickerson (8) estimated with a mutation rate of 2×10^{-8} per base pair per generation and a human genome of over 3 billion base pairs, each genome carries 60 new mutations per generation. Sun et al. (9) estimated a mutation rate of 1.4×10^{-8} which would give a mutation rate of about 40 new mutations per generation. The world population of 7 billion has about 300 to 400 billion new mutations in the current generation. The genetic diversity of mankind is exemplified by the observation that the exons (protein coding regions) of each individual genome, referred to as the exome, encompasses ~13,000 nonsynonymous and ~7,000 potentially functional variants, posing considerable challenges in identification of disease causing DNA sequence variants (DSVs) (10,11). Despite the sequence of the human genome being 99.5% identical, the remaining 0.5% is more than adequate to provide each of us a unique genome that until sequenced will have many hidden surprises. Current knowledge indicates there are 3 million SNPs per genome, which account for over 80% of human phenotype variation, whether it is the color of one's eyes or the susceptibility to disease (12).

The Search for Disease Related Genes

A major goal is to identify DNA regions that predispose or cause cardiovascular disease. This refers to the ongoing studies that correlate physical or biochemical features (phenotype) to that of the genotype. Defining the phenotype precisely is fundamental to the discovery of the associated or causal genotype. The role of the clinician in detecting the phenotype has been crucial to this pursuit and will continue to be even more so as we further refine and specify subphenotypes. DNA can be obtained from the blood, other body fluids such as saliva, or body tissue. The approach to identify the causal genes and variants has evolved dramatically over the past 3 decades. The conventional approach of genetic linkage analysis in large families, which was very successful in linking causal DNA mutations to rare single gene disorders, has all but been replaced with the newer approaches of genome-wide association studies (GWAS) and next generation DNA sequencing (NGS) in small families and individual cases. The newer approaches not only have partially overcome a major limitation of genetic linkage in identifying the causal variant in small size families but also have afforded the opportunity to identify the causal alleles in sporadic cases with single gene diseases and the susceptibility (risk) alleles in those with the complex phenotypes.

Single Gene Disorders: The Success of Genetic Linkage Analysis

Single gene disorders are the phenotypic consequences of rare DSVs that impart large effect sizes. The mutation is both necessary and sufficient to induce the disease. Familial hypertrophic cardiomyopathy was the first cardiovascular single gene disorder for which the responsible mutation was discovered. The responsible mutation was a missense mutation in the gene that encodes the beta-cardiac myosin heavy chain (13). Introducing the human mutant gene as a transgene induced the disease in both the mouse (14) and the rabbit (15). While the rare variant is sufficient to cause the disease, there is often variable expressivity (severity of the phenotype), determined by other genetic and nongenetic factors. The conventional approach for mapping the chromosomal location (locus) of the gene responsible for a single gene disorder has been genetic linkage analysis. In this technique, DNA of members of a 2- generation to 3-generation pedigree affected with the disease are genotyped using a few hundred short tandem repeat DNA markers. DNA markers that are inherited more commonly than by chance by the affected members of the family indicate the markers are in close physical proximity to the DNA region containing the responsible gene. Sequencing of candidate genes at the mapped locus usually identifies the causal variant. This approach has been exceedingly successful in mapping the causal genes for various single gene disorders, typically in large and moderate size families. It is estimated there are about 6,000 single gene disorders of which causative genes have been discovered for over 3,500 (16). Accordingly, several dozen genes for hereditary cardiomyopathies, including dilated, hypertrophic, and arrhythmogenic cardiomyopathies; hereditary arrhythmias,