European Journal of Internal Medicine xxx (2017) xxx–xxx

Contents lists available at ScienceDirect

European Journal of Internal Medicine

journal homepage: www.elsevier.com/locate/ejim

Review Article

Genetics of common complex diseases: a view from Iceland

David O. Arnar [a,b,*], Runolfur Palsson [b,c]

[a] Division of Cardiology, Internal Medicine Services, Landspitali – The National University Hospital of Iceland, Reykjavik, Iceland
[b] Faculty of Medicine, School of Health Sciences, University of Iceland, Reykjavik, Iceland
[c] Division of Nephrology, Internal Medicine Services, Landspitali – The National University Hospital of Iceland, Reykjavik, Iceland

ARTICLE INFO

Article history:
Received 27 February 2017
Received in revised form 22 March 2017
Accepted 24 March 2017
Available online xxxx

Keywords:
Genealogy
Genetics
Genomics
Genotyping
Genome-wide association studies
Whole-genome sequencing
Precision medicine

ABSTRACT

In the past decade, large scale genotyping has led to discoveries of numerous sequence variants that confer increased risk of many common complex diseases. Interestingly, a substantial proportion of pioneering genetic work has originated from the small nation of Iceland and has been facilitated by an extensive genealogy database. We provide examples of relevant observations made so far in several major disease categories central to internal medicine practice. Some of these findings offer new mechanistic clues into the pathophysiology of common disorders and may suggest novel approaches in diagnosis and drug therapy. However, a number of unresolved issues remain that will be subject of future research, driven by recent advances in high-throughput sequencing of the genome. At the same time, we are ready to begin transforming the abundant existing genetic data into practical clinical knowledge with the aim of improving the delivery of medical care. The era of precision medicine has arrived.

1. Introduction

Common complex diseases are caused by the interaction between multiple genetic factors and environmental exposures. Over the past decade, major advances in human genetics have led to discoveries of a large number of sequence variants in DNA that associate with increased risk of more than a hundred common complex diseases. This exciting era was heralded by the completion of the Human Genome Project in 2003 [1] and the International HapMap Project in 2005 [2], which provided effective instruments for study of genetic contributions to common diseases. These tools include public databases containing the reference human genome sequence, a map of human genetic variation and new technologies that allow quick and accurate analysis of whole-genome samples. It was the deciphering of the haplotypic structure of the human genome that made it possible to survey for common genetic variants by genotyping hundreds of thousands of single nucleotide polymorphisms (SNPs) simultaneously in large groups of individuals. The development of microarray methods for rapid genotyping greatly facilitated studies of associations of common sequence variants with complex diseases. However, these techniques are less effective for genotyping structural variants, such as insertions, deletions, inversions, and copy number variants (CNVs), which are abundant in the human genome, though not as common as SNPs.

Genome-wide association studies (GWAS) have yielded >10,000 published associations between DNA sequence variants and human diseases and traits [3]. In recent years, the advent of high-throughput sequencing of the genome has enabled identification of low-frequency and rare variants in the DNA sequence, thereby further advancing our ability to study the role of genetic factors in the pathogenesis of common complex diseases [4]. The transition from GWAS based on common SNPs on microarrays to studies analyzing a vast number of rare variants detected by whole-genome sequencing (WGS) and whole-exome sequencing (WES) presents new opportunities and challenges alike. At the beginning of the genetic revolution, it was widely believed that rapid changes in medical practice would soon occur as enhanced understanding of the pathogenesis of many chronic diseases would lead to more effective and focused therapies which could be tailored to each individual. We are still far away from achieving this goal, although genetic research has nevertheless led to accumulation of knowledge and better understanding of the role of the human genome in health and disease.

2. Genome-wide association studies

The first successful GWAS, published in Science in 2005, reported two SNPs with significantly altered allele frequency in patients with age-related macular degeneration [5]. Since then, GWAS have revealed numerous common sequence variants associated with a wide range of disorders frequently cared for by internists, such as coronary artery disease (CAD), atrial fibrillation (AF), type 2 diabetes (T2D), chronic kidney disease, dementia and cancer. The principle behind GWAS was to test

* Corresponding author at: Division of Cardiology, Internal Medicine Services, Landspitali – The National University Hospital of Iceland, Reykjavik, Iceland.
E-mail address: davidar@landspitali.is (D.O. Arnar).

http://dx.doi.org/10.1016/j.ejim.2017.03.018
0953-6205/

hundreds of thousands or even more than one million common SNPs in individuals with a particular disease or trait and compare the allele distribution with control subjects. Approximately 500,000 carefully selected SNP's cover >80% of common variation in populations of European ancestry [6]. The common disease-common variant hypothesis assumes that the genetic impact on common complex diseases is attributable to a limited number of common allelic variants present in >5% of the population. Minor allele frequency (MAF) refers to the frequency at which the second most common allele occurs in a given population. It is widely used in genetic studies because it is useful for distinguishing common and rare variants in the population.

Stringent methodology is essential for correct interpretation of GWAS data. Robust significance cut-off ($P < 5 \times 10^{-8}$) is necessary to avoid false positive results, due to multiple hypothesis testing (Bonferoni method) [7]. However, one must bear in mind that important SNP's not reaching this level of significance could potentially be missed. Very large sample sizes are frequently required to identify a SNP association with genome-wide significance. Moreover, replication of the initial discovery cohort in a separate sample of cases and controls is generally considered mandatory.

Most of the common sequence variants detected by GWAS in various complex diseases are associated with very modest increase in risk, usually in the range of 10–40%. Age-related macular degeneration is the best example of a common complex disease, in which the heritability is largely explained by a small number of common variants with large effect sizes [8]. By contrast, a large fraction of the heritability remains unaccounted for in most complex diseases. Nonetheless, even small to modest genetic effects can provide insight into the pathophysiology of complex diseases. In fact, the pathobiological significance of common disease alleles appears to result from multiple sequence variants with small effects and their interactions. Thus, GWAS have led to substantial advances in understanding the variation in the genomic susceptibility to disease states. However, it should be noted that GWAS are associated with important limitations. In addition to the small or modest effect sizes, the elucidation of disease mechanisms has been hampered by many associated sequence variants falling outside of the coding regions, possibly suggesting a role in gene regulation [9]. Indeed, causal variants have rarely been identified but rather susceptibility loci or genomic regions that contain one or more variants affecting disease risk. Finally, phenotypic diversity has also proved greater than initially thought.

3. Whole-genome and whole-exome sequencing

The GWAS era has taught us that studies of the human genome are much more complex than initially thought. Common variants with large effect sizes are very rare and rare variants with large effects are usually associated with Mendelian disorders (Fig. 1). Therefore, it is unlikely that rare variants with large effects contribute significantly to the heritability of most complex traits. On the other hand, it appears likely that a number of rare (MAF <0.5%) or low-frequency (MAF 0.5–5%) variants with low to modest effects contribute substantially to the heritability of complex diseases, although this may be hard to detect because of power issues. The recent transition to studies based on WGS and WES provides exciting opportunities to further explore the relationship between rare sequence variants and common diseases. Whole-genome sequencing is the most comprehensive method for analyzing the genome and reveals the complete DNA architecture [10]. The cost has rapidly decreased since this method first became available and is now only a fraction of the original price. Whole-genome sequencing is expected to yield rare variants with moderate effects and CNVs which GWAS may have missed. High-throughput sequencing is extremely efficient in Mendelian disorders, while results in complex diseases may be more variable.

Several studies based on WGS and WES have already uncovered rare variants associated with common complex diseases, including AF and sick sinus syndrome [11,12]. These studies also provide unprecedented information about human sequence diversity and insights into the ancestry of human populations.

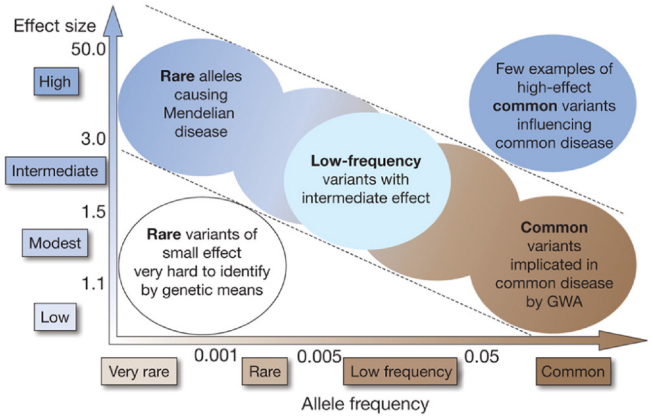

Fig. 1. Feasibility of identifying genetic risk variants by risk allele frequency and strengths of the genetic effects (odds ratio). From Manolio T, et al., Nature 2009;461:747–753, with permission.

4. Contribution of Iceland to genetic research

While many countries have contributed extensively to the recent advances in human genetics and genomics, it is intriguing that a considerable proportion of the pioneering work has originated from the small nation of Iceland. This can largely be attributed to research carried out at deCODE genetics, a human genetic research institute founded in 1996 by Dr. Kari Stefansson. From the outset, the principal goal of deCODE genetics was to find genetic variation in humans that associates with common complex diseases and traits and apply this knowledge to uncover pathogenic mechanisms to guide the development of diagnostic tests and drug therapies. To date, deCODE in collaboration with the University Hospital in Reykjavik and several other Icelandic institutions, as well as investigators worldwide, has discovered numerous genes believed to be involved in a host of complex diseases, including cardiovascular disease, diabetes, chronic kidney disease, kidney stones, cancer, Alzheimer's disease (AD) and schizophrenia (http://www.decode.is).

Despite a population of only 330.000, which largely remained isolated for many centuries, inbreeding has been shown to be modest, a finding that is also reflected by a low incidence of autosomal recessive disorders [13]. The population is relatively homogenous and the genetic makeup is a Northern European mix, with approximately 75% of the males of Scandinavian origin and 66% of the women of Celtic origin [14]. In addition, an extensive genealogy database in Iceland has greatly facilitated research in this field. This database, which is derived from church books, contains information about almost all ancestors of contemporary Icelanders dating back to the year 1650.

The majority of Icelanders have participated in the research projects at deCODE. Until now, about 150.000 Icelanders have been genotyped using DNA microarray platforms and close to 40.000 of those individuals have also had their entire genome sequenced (Stefansson K, personal communication). The large fraction of the Icelandic population that has been genotyped, allows for precise long-range phasing of the genome, thereby enabling accurate imputation of variants obtained by whole-genome sequencing into the chip-typed individuals [11]. Imputation denotes prediction of unobserved sequence variants, which can be accurately carried out to an allelic frequency of 0.01%. Furthermore, the extensive knowledge of the Icelandic genealogy permits reliable imputation of genotypes into close relatives of chip-typed individuals, creating in silico genotypes for virtually all Icelanders. The large set of imputed genomes can then be tested for association with an extensive range of phenotypes. The availability of genetic information for the

Please cite this article as: Arnar DO, Palsson R, Genetics of common complex diseases: a view from Iceland, Eur J Intern Med (2017), http://dx.doi.org/10.1016/j.ejim.2017.03.018