

Genetics of complex diseases: knowing gene polymorphisms do matter

Genética de doenças complexas: conhecer polimorfismos gênicos importa

Genética de las enfermedades complejas: importante conocer los polimorfismos genéticos

Milton Ozório Moraes ¹
Antonio Guilherme Pacheco ²

¹ Laboratório de Hanseníase,
Instituto Oswaldo Cruz,
Fundação Oswaldo Cruz,
Rio de Janeiro, Brasil.
² Programa de Computação
Científica, Fundação
Oswaldo Cruz, Rio de Janeiro,
Brasil.

Correspondence

M. O. Moraes
Laboratório de Hanseníase
Instituto Oswaldo Cruz,
Fundação Oswaldo Cruz,
Av. Brasil 4365, Rio de
Janeiro, RJ 21040-360, Brasil.
mmoraes@fiocruz.br

At the beginning of the 1990s, a very ambitious project was launched to fully map the human genome. At that time, researchers were excited with the possibility of unraveling the most basic nature of human beings and several books and papers were written emphasizing the advantages and the wonder in uncovering the human “blueprint”. The media published hopeful stories about the possibility that this knowledge could help to find the cure for every disease. At the same time a more multifaceted view implying that diseases are composite phenotypes that demand an intimate interaction of genomes with the environment is more likely. For complex diseases, hundreds of interconnected genes and environmental factors are responsible for triggering disease states. In the year 2000, the first draft of the human genome project was published after ten years of careful mapping and organization of nucleotide sequences from each of the 22 autosomal and also sexual (X,Y) human chromosomes ¹. The announcement of the complete genome was met with great expectations. However, after a few years, excitement was replaced with disillusion, since rapid solutions could not be achieved. In fact, the inability of the full genome map to provide immediate answers about the mechanisms of most diseases, such as infectious, autoimmune and cardiovascular diseases, as well as cancer, among others was frustrating.

Nevertheless, geneticists and molecular biologists began to improve the quality of the information that was formally obtained with the draft of the human genome. It is important to note that in the past 10 years, since the original genome draft, cheaper and faster sequencing and PCR-driven technologies also became available facilitating different types of genomic studies: from the deepening of the map of the human variation (for example, Hapmap, 1.000 genomes projects) to a better understanding of the molecular switches that turns genes on and off (ENCODE). But it was only more recently that the genome investments have begun to pay off.

The sequencing completion of several genomes associated with hundreds of projects that started to locate and estimate frequencies of different types of polymorphisms in thousands of individuals allowed a more thorough analysis of human variation across different populations. Case-control studies have become a popular epidemiological approach to test genetic associations with diseases, because they allow the use of most informative single nucleotide polymorphisms (SNPs), which are biallelic, spread throughout all chromosomes with a reasonable density and easy to genotype. In this case, the classic approach is to look for SNPs in candidate genes, i.e. a specific locus that has been picked-up due to some previous information in the

studies of that disease. But powerful techniques are now available, and one million SNPs from an individual can easily be genotyped, while costs for this whole genome variation analysis are decreasing fast. These are called genome-wide association studies (GWAS) with case-control designs and have been broadly employed to capture genetic information on complex diseases. In this case candidate genes are coming out after this entire genome screening. Also, genetic and epidemiology are developing novel and refined statistical tools to screen these results. Nevertheless, the great challenge is to clearly define a specific and well characterized phenotype in order to screen the genetic variation associated with it. A single feature that can be tested using case-control designs can depict the associated allele. Obviously, it is not always easy to define a well characterized phenotype since the complexity of the feature may be cryptic and could not be broken down; besides, the genotype-phenotype relationship itself may be more complex, involving other genetic and non-genetic factors that add up to a set of sufficient causes.

Nevertheless, some examples can be used to illustrate the current applications of genomic research in health, where knowing specific genotypes help predict and interfere with the phenotype. Recently in the case of the hepatitis C virus (HCV), chronic-infected patients who were unable to respond to interferon-alpha (IFN- α) treatment combined with ribavirin were tested using GWAS, which is a relatively simple approach that can be considered as a hypothesis generating method that will be later tested by more specific designs. In this case, HCV patients were grouped as to whether or not they exhibited a sustained virological response (SVR), which means patients were able to clear HCV six months after the interruption of treatment. The complete genome was searched for all polymorphisms in both groups and elegant statistical analyses were applied. A SNP located in the interleukin-28B (*IL28B*) gene was able to predict the response to IFN- α . Individuals carrying the TT genotype at the rs12979860 SNP were shown to be poor responders and results showed impressive high odds-ratio values in different ethnicities². In other populations, different SNPs spanning the same locus, in the *IL28B* gene, were associated with the SVR. To test these hypotheses more specifically, several case-control studies in dozens of different populations were conducted and eventually corroborated the association of *IL28B* SNPs and SVR. One important genetic premise was maintained when evaluating this findings: several different genotypes (in this case based on single nucleotide variations) can generate the same phenotype. Then,

it was found out that the *IL28B*, a type III interferon, played a crucial role in activating antiviral responses. So far, a direct genotype-phenotype correlation demonstrating how the SNP is finally regulating IL-28B levels that leads to HCV clearance has not yet been established.

The other astonishing genome finding was the association of specific polymorphisms with very low LDL (low density lipoprotein) levels. The levels of the so called "bad" cholesterol are regulated by another gene called pro-protein convertase subtilisin/kexin type 9 (PCSK9). Actually, PCSK9 mediates degradation of the LDL receptor (LDLR). Thus, gain-of-function mutations in PCSK9 induce degradation of LDLR that increases the circulating levels of LDL. On the other hand, a loss-of-function polymorphism at PCSK9 does not interfere with LDLR that leads to increased uptake of cholesterol, in turn leading to low levels of circulating LDL³. As a consequence, these SNPs have also been associated with decreased risk of coronary heart disease. It has been reported that rare polymorphisms in two women showed lower levels of LDL, although no clinical conditions were found. In view of that, an antibody targeting PCSK9 could reduce its circulating levels that consequently would decrease LDL drastically. Different biopharmas are now testing this approach as a highly specific treatment based on the discovery of a genetic polymorphism. Again, a very specific phenotype, LDL levels, was chosen though it was not yet apparent whether this genetic mechanism is indeed a model that could be found in other conditions or if it was an exception.

Another example confirms this idea. In multiple sclerosis it has been observed that a drug (anti-TNF) can mimic a genetic effect. A GWAS screening revealed a SNP (rs 1800693), in the tumor necrosis factor receptor (*TNFR*) gene, that was associated with disease susceptibility. This genetic variation was responsible for a rare soluble form of TNFR that blocks circulating TNF⁴ and has been suggested that TNF blockage could trigger MS. Curiously, anti-TNF drugs (such as infliximab) largely used for other inflammatory diseases such as rheumatoid arthritis, psoriasis and Crohn's disease can also trigger multiple sclerosis indicating clearly that genome-based analysis unveiling SNPs are being associated with different complex diseases and their mechanisms of action can be copied using different types of immunobiologicals or drugs.

It is obvious, though, that good genetic experimental designs including large samples (several independent replications involving thousands of people are expected) and well characterized groups with clear phenotypes are needed to dis-

entangle the genetic circuitry. These intricate mechanisms also need a careful analysis and functional validation, which means that it is very important to analyze genetic findings in light of the role that genetic variation has in the regulation of its products. Unfortunately, some diseases are quite difficult to define well characterized phenotypes. Sometimes, the phenotype “disease per se” is a combination of different features and GWAS analysis is not able to depict a number of candidate genes. For example, in tuberculosis af-

ter a large screening only one SNP in a poor gene region was replicated⁵. Thus, “disease-wide” phenotypes could be biased leading to spurious associations and should be avoided. Again, a careful definition of the phenotype should be prioritized.

In summary, recent advances are promising, but to detect these specific genotype-phenotype associations throughout the genome is still very challenging, and careful research with good experimental designs are needed.

Contributors

Both authors drafted and revised the text.

References

1. Collins FS. Genome-sequencing anniversary. Facets of the genome. *Science* 2011; 331:546.
2. Gonzalez SA, Keeffe EB. IL-28B As a predictor of sustained virologic response in patients with chronic Hepatitis C virus infection. *Gastroenterol Hepatol (NY)* 2011; 7:366-73.
3. Cohen JC, Hobbs HH. Genetics. Simple genetics for a complex disease. *Science* 2013; 340:689-90.
4. Gregory AP, Dendrou CA, Attfield KE, Haghikia A, Xifara DK, Butter E, et al. TNF receptor 1 genetic risk mirrors outcome of anti-TNF therapy in multiple sclerosis. *Nature* 2012; 488:508-11.
5. Thye T, Vannberg FO, Wong SH, Owusu-Dabo E, Osei I, Gyapong J, et al. Genome-wide association analyses identifies a susceptibility locus for tuberculosis on chromosome 18q11.2. *Nat Genet* 42:739-41.

Submitted on 06/Aug/2013

Approved on 13/Aug/2013