Reumatología Clínica

Editorial

Genome-Wide Association Studies (GWAS) in Complex Diseases: Advantages and Limitations☆

Enfermedades complejas y análisis genéticos por el método GWAS. Ventajas y limitaciones

José A. Riancho

Departamento de Medicina Interna, Hospital Universitario Marqués de Valdecilla-IFIMAV, Universidad de Cantabria, Santander, Spain

Classic inherited diseases are caused by a single gene mutation, often with serious consequences for the organism, but are fortunately rare. Acquired diseases, on the contrary, are due to environmental factors. However, many of the most prevalent diseases are actually the result of the combination of hereditary and environmental factors. Many common disorders such as osteoporosis, arthritis, diabetes or hypertension tend to cluster in families, reflecting their hereditary component (although there may also be shared environmental factors). The importance of hereditary factors in osteoporosis is large and, for example, accounts for between 50% and 80% of the variability of bone mass.[1] Genetic epidemiology studies showed that, unlike classical hereditary diseases, the risk for these disorders is not explained by the alteration of a single gene. Hence the name of "polygenic" or "complex" diseases.

By sequencing the human genome it was ascertained that there are many inter-individual variations in DNA. Contrary to mutations, these variations are quite common and in many cases their functional impact is limited (in fact, most occur in non-coding regions) and were called "polymorphisms". Among them, the single nucleotide polymorphisms (SNPs) consist of simply the change of one base for another. They are very frequent, about 15 million in the genome. There are also frequent repeat polymorphisms, which consist of groups of a few nucleotides that are repeated variable number of times in individuals. More recently, another form of DNA variation was identified and called "variation in the number of copies." It consists of relatively large regions of the genome, thousands of nucleotides, which in some individuals are repeated on the same or on different chromosomes.

The fact that complex diseases do not follow a classic pattern of inheritance, i.e. they cannot be explained by a single gene disorder, along with the discovery of the frequency of polymorphisms led to the hypothesis of "common illnesses common variants". This assumes that common complex diseases are the result of the combined effect of many common polymorphisms in the population.

After the epidemiological studies demonstrated the important genetic component of these disorders, the researchers set out to try to identify the genes and polymorphisms involved. To do this, based on knowledge of the biology and pathogenesis of disease, "candidate genes" were identified and its association with these disorders explored. For example, given the important role of vitamin D and sex hormones on bone metabolism, some of the first candidate gene studies examined the association of polymorphisms of the vitamin D and estrogen receptors with osteoporosis.[2,3]

While their design allows multiple variations, these studies in essence choose a candidate gene, identify some of its polymorphisms and examine whether polymorphic alleles at these loci are associated with a particular phenotypic trait or frequency of a disease. Subsequently, there have been many studies on other candidate genes in relation to a variety of disorders, but overall they have not responded to the expectations generated by the attractive hypothesis that sustained them. The results obtained by some researchers are often not replicated in other studies and the strength of association between genotype and phenotype has generally been small.[4]

Further development of microarrays made it possible to analyze hundreds of thousands of SNPs in an efficient manner, with a small DNA sample and a much lower cost than would be needed to study SNP individually. Furthermore, since these SNPs distributed throughout the chromosomes, a new approach was possible: to explore the whole genome without prior hypothesis, i.e. without previously selecting candidate genes. SNPs included in the microarray were selected considering the patterns of linkage, so that other polymorphisms that are not directly related are also captured. These genome-wide association studies (GWAS) raised high expectations. It was thought that they would finally allow identifying all the genes in the heritability of complex diseases. In fact, since 2005, some 1200 GWAS SNP associations have been published with more than 200 diseases or phenotypic traits described (you can find a listing on www.genome.gov/GWAStudies).

However, the results of GWAS have not responded to initial expectations. Many genes found associated with a particular disease have no known biological effects to explain this relationship. However, this may not be but a reflection of the incompleteness of our knowledge. In fact, those findings are being used to identify new pathogenic mechanisms and new therapeutic targets.[5] More

---

surprising is the fact that, even when combining all available GWAS on a particular disorder, polymorphisms usually associated explain less than 5%–10% of the risk of disease.

How is it possible that studies that analyzed 500 000 SNPs in several thousand individuals shed these poor results in terms of risk prediction? One reason may be the power of the studies. By exploring many SNPs there is a high risk of finding false significant associations according to conventional criteria of $P<0.05$. In fact, when 500 000 SNPs were analyzed, one would expect to find over 25 000 diseases associated with a P-value <.05 simply by chance. To avoid this error, a correction for multiple comparisons is employed, so that associations are considered significant with a P less than 10^{-7} or 10^{-8}. This approach reduces false positives, but also markedly decreases the power to detect SNPs associated with disease.[6] To increase power one can only increase the sample size, or use larger studies, which is increasingly common, using meta-analysis of several studies. However, it is estimated that GWAS have been identified and presumably more than 80%–90% of common SNPs are associated with prevalent disorders such as osteoporosis, with an odds ratio greater than 1.1–1.2. Therefore, we think that if studies are extended, it is likely that other SNPs are also associated with disease, but the individual influence of each of them will presumably be very small.[7,8]

So what determines that much of the risk of disease still remains unexplained? The answer to this question is uncertain. One possibility is that the prevalent diseases are not the result of common variants with relatively small individual effects (such as assuming the hypothesis of common diseases, common variants), but rare variants with relatively large effects are not identified in GWAS. Keep in mind that, by design, the microarrays used in GWAS are unable to detect the influence of SNPs with rare alleles with populations that have frequencies of less than 1%–10%. It is also possible that differences in risk are due to the interaction between different SNPs or between SNPs and environmental factors. Neither individual GWAS studies nor the meta-analyses conducted so far have enough power to unravel these interactions. We cannot exclude that other forms of genetic variation, such as repeat polymorphisms or variations in the number of copies, which have hardly been explored, play an important role. Possibly epigenetic mechanisms (which are potentially heritable and can modulate the expression of genes without involving changes in DNA sequence) also significantly influence the risk of disease. It is also possible that there is a high genetic heterogeneity in the pathogenesis of these diseases. Indeed, the combination of several studies using meta-analysis increases power to detect genetic factors common to all the populations studied. But that strategy is not necessarily effective when combining studies with groups of individuals in which the influential genes are different. For example, polymorphisms of the aromatase enzyme that converts androgenic precursors into estrogens in peripheral tissues are associated with bone mass in postmenopausal women, but not in young women with active ovarian estrogen production.[9] Clearly, this association may be masked in a meta-analysis with mixed pre-and postmenopausal women.

In short, GWAS represent a huge technological advance that has identified new genes associated with various diseases. This offers interesting possibilities for the development of new treatments, but so far the results have been disappointing in predicting the overall risk of disease. Thus, researchers in this field have before finding out what this still unknown 'dark matter' explains regarding the heritability of complex diseases.

Disclosure

The authors have no disclosures to make.

References

1. Ralston SH. Osteoporosis as an hereditary disease. Clin Rev Bone Miner Metab. 2010;8:68–76.
2. Morrison NA, Qi JC, Tokita A, Kelly PA, Crofts L, Nguyen TV, et al. Prediction of bone density from vitamin D receptor alleles. Nature. 1994;367:284–7.
3. Kobayashi N, Fujino T, Shirogane T, Furuta I, Kobamatsu Y, Yaegashi M, et al. Estrogen receptor alpha polymorphism as a genetic marker for bone loss, vertebral fractures and susceptibility to estrogen. Maturitas. 2002;41:193–201.
4. Richards JB, Kavvoura FK, Rivadeneira F, Styrkarsdottir U, Estrada K, Halldorsson BV, et al. Collaborative meta-analysis: associations of 150 candidate genes with osteoporosis and osteoporotic fracture. Ann Intern Med. 2009;151:528–37.
5. Bossini-castillo L, Martin JE, Diaz-Gallo LM, Rueda B, Martin J. Genética de la esclerodermia. Reumatol Clin. 2010;6:12–5.
6. Panoutsopoulou K, Southam L, Elliott KS, Wrayner N, Zhai G, Beazley C, et al. Insights into the genetic architecture of osteoarthritis from stage 1 of the arcOGEN study. Ann Rheum Dis. 2011;70:864–7.
7. Duncan EL, Brown MA. Genetic determinants of bone density and fracture risk–state of the art and future directions. J Clin Endocrinol Metab. 2010;95:2576–87.
8. Goldstein DB. Common genetic variation and human traits. N Engl J Med. 2009;360:1696–8.
9. Zarrabeitia MT, Hernandez JL, Valero C, Zarrabeitia AL, Garcia-Unzueta MT, Amado JA, et al. A common polymorphism in the 5′-untranslated region of the aromatase gene influences bone mass and fracture risk. Eur J Endocrinol. 2004;150:699–704.