

The role of population genetics and its implications in the study of complex-Disease risk.

Rocío Gómez*

Department of Toxicology, Center for Research and Advanced Studies of the National Polytechnic Institute, Mexico

Accepted on October 04, 2017

Commentary

Genetic-association is a prominent word in the biomedical databases appearing in more than one million of publications. This keyword is one of the most used to find phenotype-genotype associations, identifying risk variants that could be used as biomarkers in complex diseases. Nonetheless, this word does not necessarily imply a real association, given that the human populations, as other ones, are dynamic and consequently, an important rate of statistical errors is generated. This situation is not only reflected in the classical case-control genetic association studies. Certainly, the genome-wide association studies, as well as the whole exome sequence, are susceptible of false associations. Latino/Hispanic and African-American populations present a complex genetic architecture due to the disparities in the ancestral background of populations from which they emerged. Ancestry informative markers, Bayesian inferences of a number of subpopulations, hierarchical clustering, and ancestry matching are indispensable tools to reduce the bias caused by the population heterogeneity and the inflated findings. Herein, is a short commentary where a general overview is presented about population genetic parameters and the role of ancestry to identify risk loci and its implications for medical practice. Also, the hint of the consequences of the depreciation of these arguments was done evidencing its consequences.

Gene-disease association studies have evidenced the loci contribution in the development of complex disease being a cornerstone in the field of genetics, identifying risk biomarkers and even individual risk profiles. From the classical case-control studies, where gene frequencies between healthy and non-healthy groups are compared, to next-generation sequence (NGS) strategies, the aim is to identify susceptibility biomarkers to medicine practice. In the last decade, genome-wide association studies (GWAS) have significantly contributed to the association of variants with the disease phenotypes. More recently, whole-exome sequencing (WES) strategies have identified rare variants responsible for genetic risk of different diseases. Nevertheless, the results have been discrepant among populations. To avoid spurious associations, statistical models have been implemented trying to control the false discovery rates and to reinforce the good publication practices. Overall, these tools assume a homogeneous ancestral background where individuals assort randomly. However, one of the most prominent confounding factors in phenotype-genotype associations is the genetic architecture, particularly relevant in ethnically diverse populations such as Latinos/Hispanos and African-Americans [1]. Admixture populations are a despaired

amalgam of parental populations historically divergent where different strata present dissimilar ancestry proportions and ergo different allele frequency patterns. Consequently, the genomes of individuals within an admixture population are not homogeneous being modified by sex, socio-economic status, and education level [2]. Likewise, and given that the youth of admixed populations (10-15 generations) [1], the different subpopulation conforming the whole population are not equally represented, exhibiting high rates of spurious results. However, the false-positive results can be detected through different tests based on population genetic statistical models. A guide to increasing the quality control of genetic association studies is the Strengthening the Reporting of Genetic Association studies (STREGA) [3,4] which is even a useful tool in NGS [5]. STREGA statements suggest the determination of genetic structure and the Hardy-Weinberg equilibrium (HWE), among other 22 parameters, as well as some strategies to adjust for confounding effects related to the genetic architecture of studied population [4]. HWE, an essential statistical model to genetic data analysis, reinforces data quality decreasing the skew caused by different population conditions. This model assumes random mating, and therefore the departures to this hypothesis can reflect different situations such as gene flow, assortative mating, and population stratification. All these situations caused Hardy-Weinberg departures (HWD) mainly linked to heterozygous excess ($F_{is} < 0$). This particular characteristic has also been related to genotyping errors associated with selection bias. In addition, HWD shed light on problems due to uncover or low-coverage areas, an excess of tandem repeat polymorphisms, and segmental duplications, besides to detect long-range haplotypes [5]. Conversely, homozygous excess ($F_{is} > 0$) is associated with inbreeding populations as well as null alleles existence [5]. Thus, this elementary equation indicates that our data, and consequently our findings, are highly susceptible to misinterpretation due to type I and type II statistical errors [6]. HWD could also be a signature of phenotype-genotype association related to an overrepresentation of the risk allele in the non-healthy group; the depreciation of this parameter can make weak the evidence. On the other hand, the ancestry variation among sub-populations that conform an admixed one can also provoke spurious associations because the ancestral background and the disease risk are no equilibrated among populations [7]. Thus, the "association" could be related to the ancestry rather than the disease, inflating the results even in gene-environmental studies [8]. Different strategies such as WES, GWAS and the pharmacogenomics assay, advise the inclusion of extreme phenotypes, the meticulous ancestry matching in cases and controls, and admixture correction to

adjust the dosage [9,10]. Currently, the multi-ethnicity inside admixed populations has become a relevant tool to trace back the ancestral origin of some complex diseases. Ultimately, the linkage disequilibrium patterns (LD, a non-random association of allelic states across loci), are also affected by the genetic stratification provoking a bogus LD even between loci of different chromosomes [11]. As mentioned before, the admixed populations exhibit a limited number of generations, presenting a reduced effective population and even identity-by- descend [12]. Hence, admixture LD reflects the demographic and population histories, and even inbreeding practices causing false positive rates. In this setting, assortative mating practices could cause systematic differences in the ancestral background perturbing the gene-disease associations [13]. Of note, the Latino admixture is sex-biased given the miscegenation process occurred between Amerindian women and European or African men with ensuing regional disparity [2]. As in HWD, the LD is a signal of possible association where a genomic locus could be influencing the phenotype. Therefore, the discrepant in LD patterns across ethnicities could increase gene-mapping resolution [14]. Hence, despite the “inconvenience” of population stratification, the multi-ethnicity of the admixed population may constitute a crucial issue to relate the disease risk and the ancestry [15]. In conclusion, the pinpoint knowledge of population genetics tools is a highlighted task in population-based genetic association studies.

References

1. Johnson NA. Ancestral components of admixed genomes in a Mexican cohort. *PLoS Genet.* 2011;7: e1002410.
2. Bryc K. Colloquium paper: Genome-wide patterns of population structure and admixture among Hispanic/Latino populations. *Proc Natl Acad Sci USA.* 2010;107:8954-8961.
3. Nedovic D, Panic N, Pastorino R, et al. Evaluation of the Endorsement of the Strengthening the Reporting of Genetic Association Studies (STREGA) statement on the reporting quality of published genetic association studies. *J Epidemiol.* 2016;26:399-404.
4. Little J. Strengthening the Reporting of Genetic Association Studies (STREGA)-an extension of the STROBE statement. *Genetic Epidemiol* 2009;33:581-598.
5. Graffelman J, Jain D, Weir B. A genome-wide study of Hardy-Weinberg equilibrium with next-generation sequence data. *Hum Genet.* 2017;136:727-741.
6. Namipashaki A, Razaghi-Moghadam Z, Ansari-Pour N. The essentiality of reporting Hardy-Weinberg equilibrium calculations in population-based genetic association studies. *Cell J.* 2015;17:187-192.
7. Magi R. Trans-ethnic meta-regression of genome-wide association studies accounting for ancestry increases power for discovery and improves fine-mapping resolution. *Hum Mol Genet.* 2017;26:3639-3650.
8. Hoyos-Giraldo L. The effect of genetic admixture in an association study: genetic polymorphisms and chromosome aberrations in a Colombian population exposed to organic solvents. *Ann Hum Genet.* 2013;77:308-20.
9. Do R, Kathiresan S, Abecasis GR. Exome sequencing and complex disease: Practical aspects of rare variant association studies. *Hum Mol Genet.* 2012;21:R1-R9.
10. Duconge J. A novel admixture-based pharmacogenetic approach to refine warfarin dosing in Caribbean hispanics. *PloS One.* 2016;11:e0145480.
11. Baird SJ. Exploring linkage disequilibrium. *Mol Ecol Resour.* 2015;15:1017-19.
12. Li M, Li C. Assessing departure from Hardy-Weinberg equilibrium in the presence of disease association. *Genet Epidemiol.* 2008;32:589-99.
13. Zaitlen N. The effects of migration and assortative mating on admixture linkage disequilibrium. *Genetics.* 2017;205:375-83.
14. Benyamin B. Cross-ethnic meta-analysis identifies the association of the GPX3-TNIP1 locus with amyotrophic lateral sclerosis. *Nat Commun.* 2017;8:611.
15. Chakraborty R, Weiss KM. Admixture as a tool for finding linked genes and detecting that difference from allelic association between loci. *Proc Natl Acad Sci USA.* 1988;85:9119-23.

*Correspondence to:

Rocío Gómez
Department of Toxicology
Center for Research and Advanced Studies of the
National Polytechnic Institute.
México
Tel: +52 55 5729 6000
E-mail: mrgomez@cinvestav.mx