Title: The road to unravel gene-environment interactions on cardiovascular complex diseases

Running title: Unraveling gene-environment interactions

Authors: Roberto Elosua, MD, PhD [1,2,3]

Affiliations:

1. Cardiovascular Epidemiology and Genetics Research Group, REGICOR Group. Hospital del Mar Medical Research Institute (IMIM), Barcelona, Catalonia, Spain

2. CIBER Cardiovascular, Barcelona, Catalonia, Spain

3. Medicine Department, Medical School, University of Vic - Central University of Catalonia (UVic-UCC), Vic, Catalonia, Spain

Corresponding author:

Roberto Elosua, MD, PhD

Cardiovascular Epidemiology and Genetics – IMIM

88 Dr Aiguader, 08003 Barcelona

Catalonia, Spain

Email: relosua@imim.es

Journal subject terms: Genetic, Association Studies, Smoking, Coronary Artery Disease

Genome-wide association studies (GWAS) have contributed to the identification of genetic variants associated with coronary artery disease (CAD); a recent GWAS meta-analysis reported 66 loci significantly associated with CAD.[1] However, the heritability explained by these loci remains low, around 15%. One of the limitations of GWAS is the likelihood of false negative results related to the strict and commonly defined genome-wide significance threshold that accounts for multiple comparisons using the Bonferroni correction (p-value $< 5\times10^{-8}$). Other methods to correct for the high number of comparisons, such as false discovery rate, provide a larger number of associated loci, but the explained heritability remains around 21%.[2]

The debate about the role of genetics (nature) versus the role of environmental factors (nuture) on health has classically confronted and opposed these two influences, but the interplay between the two also contributes to the pathogenesis of complex diseases and could partially explain the missing heritability of CAD. The gene-environment interaction (G×E) adds complexity to the already complex individual roles of genes and environment in the etiology of some diseases, such as CAD. There are several reasons to study G×E interactions: helping to unravel the pathogenic mechanisms of a disease such as in phenylketonuria (i.e., two necessary causes)[3] but also in complex diseases;[4] from a public health perspective, identifying the subset of individuals in whom a preventive intervention could have the most beneficial effect; and, finally, improving the model fit or its predictive capability via the inclusion of interaction terms in multivariate models.

There are several examples of established G×E interactions, with the clearest ones related to Mendelian diseases in which both the presence of a genetic variant and the

presence of an environmental factor are necessary for the occurrence of the clinical phenotype. One of the most well-known examples is phenylketonuria,[3] a disease caused by a defect in the gene encoding the enzyme phenylalanine hydroxylase that results in the accumulation of this amino acid and intellectual disability. A phenylalanine-restricted diet minimizes or prevents most of the neuropsychological complications caused by this genetic variation. However, unraveling G×E interactions related to complex diseases is not a straightforward enterprise. In a recent report Ritz et al. presented some examples of G×E interactions and extracted some lessons from these successes.[5] In this issue of *Circulation: Precision and Genomic Medicine*, Hindy et al.[6] illustrate an interaction between smoking and a polygenic risk score including 50 SNPs associated with CAD.

In this editorial, I will address some of the main features of G×E interactions and their implications for the design and analysis of studies such as that presented by Hindy et al.

Sample size and statistical power

There are two main factors that influence the statistical power and the sample size required for the study of G×E interactions: the effect size and the quality of the assessment of environment variables.

With the exception of a few number of G×E interactions, such as the interaction between smoking and HLA genotypes in the risk of multiple sclerosis,[7] the effect sizes of G×E interactions in complex diseases are usually small. Even when the effect size is big, stratified analyses can reduce the power of the study. Therefore, well-powered

studies are required to unveil these G×E interactions,[8] and collaborative efforts such as the "Gene-Lifestyle Interactions Working Group" are necessary to systematically evaluate these interactions.[9]

One of the main problems related to G×E interaction analyses is the assessment of environment variables and the measurement error, which introduces a non-differential bias underestimating the real effect size. New methodologies and initiatives are being developed to improve this assessment, and new concepts such as the "exposome" have been proposed and accepted by the scientific community.[10] The exposome is aimed at measuring not only external exposures but also internal signatures of these exposures such as DNA methylation profiles, transcriptome signatures, or other type of biomarkers.[11]

Hindy et al. categorized the exposure to smoking in three categories (never smokers, former smokers and current smokers) as is usually done in most epidemiological studies. However, a more refined exposure to smoking including the quantity of cigarettes smoked per day, the lifetime exposure to smoking (including early life exposure to passive smoking), or the use of biomarkers could have provide additional information/power to the study.

Interaction scale

The combined effect of two exposures (gene and environment) can be assessed using an additive effect or a multiplicative effect.[12] When the outcome is a continuous trait, multivariate analysis including the interaction term explores the departure from an

additive effect. When the outcome is binary, multivariate analysis explores the departure from a multiplicative effect and, commonly, the presence of an interaction in the additive scale is not explored. Several statistics such as the relative excess risk due to interaction (RERI) have been proposed to explore for the departure from an additive effect.[13]

Hindy et al. report that the combined effect of smoking and genetic risk is less than multiplicative. They do not explore the presence of a more-than-additive combined effect. In a recent report in which we analyzed the departure from an additive effect and multiplicative effect of the genetic risk load related to classical cardiovascular risk factors on CAD, we observed that the confluence of low-density lipoprotein cholesterol- and triglyceride-related genetic risk load has an additive effect on CAD risk, whereas the interaction between low-density lipoprotein cholesterol and CAD genetic load is more than multiplicative, supporting the hazardous impact on atherosclerosis progression of the combination of inflammation and increased lipid levels.[13]

Candidate gene vs. genome-wide vs. genetic risk load approaches

The candidate gene approach is the most commonly used with several examples in the cardiovascular area including environment variables that modify the effect size of genetic variants associated with CAD[14] and also genetic variants that modify the efficacy or the side effects of some drugs (pharmacogenetic).[15] Now with the availability of genome-wide and environmental data from hundreds of thousands of individuals, an agnostic approach, without previous hypotheses, is also being implemented in new consortia and biobanks.[9] This approach based on single variants

could contribute to the understanding of the pathogenesis of the disease under study and have public health implications. One of the most well-known interactions is the modifying effect of physical activity on the association between *FTO* genetic variants and obesity, which is attenuated by ~30% in physically active compared to inactive individuals.[16]

Hindy et al. use an alternative method, defining a genetic risk score and analyzing the interaction between this genetic risk load and smoking on CAD. This approach is becoming more popular and could contribute to the identification of subgroups of the population that would obtain higher benefits from modifying their exposure to the environmental factor under study.[17]

Methodological challenges

Classical case-control studies, case-only studies, and new 2-stage study designs[18] including new statistical analysis strategies[19] are being implemented. Replication of findings in independent cohorts is a key element of genetic association studies; however, replication of G×E interactions is not always a feasible strategy, and new approaches using biological knowledge and functional validation have been proposed.[20]

In the coming years, we can expect that new G×E interaction analyses will be performed more frequently, will contribute to the discovery of new loci and new mechanisms related to complex diseases, will improve the predictive capacity of risk functions, and will identify subsets of the population that would obtain the greatest

benefit from modifying lifestyle or the exposure to environmental factors—a key goal of precision medicine.

REFERENCES

1. CARDIoGRAMplusC4D Consortium. Large-scale association analysis identifies new risk loci for coronary artery disease. *Nat Genet*. 2013;45:25–33.

2. Nelson CP, Goel A, Butterworth AS, Kanoni S, Webb TR, Marouli E, et al. Association analyses based on false discovery rate implicate new loci for coronary artery disease. *Nat Genet*. 2017;49:1385–1391.

3. Camp KM, Parisi MA, Acosta PB, Berry GT, Bilder DA, Blau N, et al. Phenylketonuria Scientific Review Conference: state of the science and future research needs. *Mol Genet Metab*. 2014;112:87–122.

4. Robinson MR, English G, Moser G, Lloyd-Jones LR, Triplett MA, Zhu Z, et al. Genotype-covariate interaction effects and the heritability of adult body mass index. *Nat Genet*. 2017;49:1174–1181.

5. Ritz BR, Chatterjee N, Garcia-Closas M, Gauderman WJ, Pierce BL, Kraft P, et al. Lessons learned from past gene-environment interaction successes. *Am J Epidemiol*. 2017;186:778–786.

6. Hindy G, Wiberg F, Almgren P, Melander O, Orho-Melander M. A polygenic risk score for coronary heart disease modifies the elevated risk by cigarette smoking for disease incidence. *Circ Genom Precis Med*. 2018;11:e001856.

7. Hedström AK, Katsoulis M, Hössjer O, Bomfim IL, Oturai A, Sondergaard HB, et al. The interaction between smoking and HLA genes in multiple sclerosis: replication and refinement. *Eur J Epidemiol*. 2017;32:909–919.

8. Robinson MR, English G, Moser G, Lloyd-Jones LR, Triplett MA, Zhu Z, et al. Genotype-covariate interaction effects and the heritability of adult body mass index. *Nat Genet*. 2017;49:1174–1181.

9. Rao DC, Sung YJ, Winkler TW, Schwander K, Borecki I, Cupples LA, et al. Multiancestry study of gene-lifestyle interactions for cardiovascular traits in 610 475 individuals from 124 cohorts: design and rationale. *Circ Cardiovasc Genet*. 2017;10:e001649.

10. Wild CP. The exposome: from concept to utility. *Int J Epidemiol*. 2012;41:24–32.

11. Joehanes R, Just AC, Marioni RE, Pilling LC, Reynolds LM, Mandaviya PR, et al. Epigenetic signatures of cigarette smoking. *Circ Cardiovasc Genet*. 2016;9:436–447.

12. VanderWeele TJ, Knol MJ. A tutorial on interaction. *Epidemiol Methods.* 2014;3:33–72.

13. Elosua R, Lluís-Ganella C, Subirana I, Havulinna A, Läll K, Lucas G, et al. Cardiovascular risk factors and ischemic heart disease: is the confluence of risk factors greater than the parts? A genetic approach. *Circ Cardiovasc Genet*. 2016;9:279–286.

14. Saleheen D, Zhao W, Young R, Nelson CP, Ho W, Ferguson JF, et al. Loss of cardioprotective effects at the ADAMTS7 locus as a result of gene-smoking interactions. *Circulation*. 2017;135:2336–2353.

15. Elosua R, Sayols-Baixeras S. The genetics of ischemic heart disease: from current knowledge to clinical implications. *Rev Esp Cardiol (Engl Ed)*. 2017;70:754–762.

16. Graff M, Scott RA, Justice AE, Young KL, Feitosa MF, Barata L, et al. Genome-wide physical activity interactions in adiposity—a meta-analysis of 200,452 adults. *PLoS Genet*. 2017;13:e1006528.

17. Aschard H, Tobin MD, Hancock DB, Skurnik D, Sood A, James A, et al. Evidence for large-scale gene-by-smoking interaction effects on pulmonary function. *Int J Epidemiol*. 2017;46:894–904.

18. Stenzel SL, Ahn J, Boonstra PS, Gruber SB, Mukherjee B. The impact of exposure-biased sampling designs on detection of gene-environment interactions in case-control studies with potential exposure misclassification. *Eur J Epidemiol*. 2015;30:413–423.

19. Gauderman WJ, Mukherjee B, Aschard H, Hsu L, Lewinger JP, Patel CJ, et al. Update on the state of the science for analytical methods for gene-environment interactions. *Am J Epidemiol*. 2017;186:762–770.

20. Ritchie MD, Davis JR, Aschard H, Battle A, Conti D, Du M, et al. Incorporation of biological knowledge into the study of gene-environment interactions. *Am J Epidemiol*. 2017;186:771–777.