Rev Esp Cardiol. 2017;70(9):696–698

Editorial

The State of Cardiovascular Genomics: Abundant Data, Limited Information

Situación actual en genómica cardiovascular: muchos datos, poca información

Stella Aslibekyan[a,*] and Edward A. Ruiz-Narváez[b]

[a] Department of Epidemiology, University of Alabama at Birmingham, Birmingham, Alabama, United States
[b] Slone Epidemiology Center, Boston University, Boston, Massachusetts, United States

Article history:
Available online 8 April 2017

"*Where is the wisdom we have lost in knowledge? Where is the knowledge we have lost in information?*"
T.S. Eliot, *Choruses from the Rock*, 1934

In the 15 years since the completion of the Human Genome Project, the field of cardiovascular genetics and genomics has undergone a remarkable transformation. From targeted, hypothesis-driven investigations of Mendelian disorders such as familial hypercholesterolemia, studies of genetic contributions to cardiovascular disease have quickly expanded to agnostic, genome-wide scans of increasing resolution, essentially becoming a 'big data' problem. The availability of whole genome-sequencing data continues to grow exponentially: in the coming years, the Trans-Omics for Precision Medicine[1] and the Centers for Common Disease Genomics[2] programs in the United States are poised to sequence more than 250 000 individual genomes, approximately, of which around 25 000 individuals will be cases of early onset coronary artery disease and stroke. In addition to institutional commitment, this data explosion was enabled by technological advances that have dramatically lowered genotyping costs, with the promise of the $100 genome now on the horizon.[3]

The first wave of large-scale genome-wide association studies (GWAS) of cardiovascular phenotypes has produced several findings that were biologically plausible, yet explained only a small proportion of the overall phenotypic variability. For example, the seminal GWAS of circulating lipids conducted in > 100 000 individuals of European ancestry identified 95 significant loci accounting for only 10% to 12% of the overall trait variance and 25% to 30% of the heritable component.[4] A subsequent effort expanding the search beyond common (minor allele frequency > 5%) variants in > 180 000 individuals uncovered associations with 62 additional loci that cumulatively enhanced the explained variance by < 3%.[5] Similarly, recent large-scale GWAS of blood pressure traits[6,7] have identified approximately 50 relevant loci explaining only about 2% of the phenotypic variance.[8] The largest meta-analysis of coronary artery disease GWAS to date revealed a similar genetic architecture, with most of the heritability (approximately 13%) explained by common loci and rare variants adding only another 2%.[9] Notably, the modest explanations offered by GWAS findings are not always indicative of limited therapeutic potential. This is especially well illustrated by the *HMGCR* polymorphisms, which exhibit small effect sizes in GWAS,[4] yet statin therapies targeting its gene product have been remarkably successful for lowering cardiovascular risk.[10] Similarly, the promise of PCSK9-targeting therapies, discussed recently in a review article published in *Revista Española de Cardiología*,[11] is disproportionate to the phenotypic variance explained (< 1% in plasma triglyceride levels).[12] Because variance explained depends on both effect size and allele frequency in the population, genetic findings that will emerge from future exome- and whole genome-sequencing studies—illustrated by the example of *APOC*[13]—could still elucidate the underlying mechanisms and inform therapeutic developments despite their limited contributions to trait heritability. To locate such valuable findings in the haystack of genome-wide data, however, it may be wise to incorporate other layers of -omics and functional data, using bioinformatic algorithms to prioritize functionally relevant variants.

The study of genome-wide epigenetic variation, specifically DNA methylation, has so far been another fruitful avenue of inquiry. Epigenetic processes embody environmental influences such as diet, lifestyle, and other factors by direct biochemical modifications of the DNA molecule. Although the extent, if any, of transgenerational epigenetic inheritance in humans is still unclear,[14,15] it may potentially add to the total heritability of any given trait. In the first published studies of plasma lipids, up to 8 methylation loci explained 5.5% to 11.6% of triglyceride variation across several cohorts[16–18]; several consortia meta-analyses of other cardiovascular phenotypes (incident coronary heart disease, circulating cytokines, hypertension, and others) are currently underway and have shown promising preliminary results.[19] Although this relative success of the epigenome-wide studies may in part be due to the 'winner's curse,'[20] it also has biological underpinnings: methylation variation, with its corresponding changes in gene transcription, is more proximal to the phenotype

SEE RELATED CONTENT:
http://dx.doi.org/10.1016/j.rec.2017.05.011, *Rev Esp Cardiol.* 2017;70:744–753.
http://dx.doi.org/10.1016/j.rec.2017.02.046, *Rev Esp Cardiol.* 2017;70:754–762.
http://dx.doi.org/10.1016/j.rec.2017.05.013, *Rev Esp Cardiol.* 2017;70:763–769.

* Corresponding author: Department of Epidemiology, University of Alabama at Birmingham, 1665 University Blvd, RPHB 230J, Birmingham, AL 35205, United States.
E-mail address: saslibek@uab.edu (S. Aslibekyan).

http://dx.doi.org/10.1016/j.rec.2017.03.012

than changes in the DNA sequence, and is thus more likely to have a larger effect. However, most methylome-wide studies conducted in large cohorts are plagued by their cross-sectional nature (which precludes from establishing temporality or causality) as well as the unavailability of the biologically relevant tissue (eg, liver for lipid metabolism). Recent attempts to address the causality challenge using Mendelian randomization techniques[21] were limited by the lack of robust genetic instruments for complex traits, such as plasma lipids or blood pressure, and hampered by the low proportion of variance explained by known polymorphisms, as discussed above. Therefore, there is pressing need for carefully controlled functional studies in vitro and in animal models to test the causality of genes identified by genome- and epigenome-wide studies of cardiovascular traits.

Epigenetic processes also play a critical role in programming cardiometabolic risk during the early stages of development.[21,22] A consistent body of evidence shows that impaired fetal growth is associated with DNA methylation levels. For example, individuals who experienced in utero exposure to famine during the Dutch Hunger Winter (1944-1945) had differential whole-blood methylation levels of the insulin-like growth factor 2 (*IGF2*) gene 6 decades later,[23] as well as in 6 out of 15 other candidate genes involved in metabolic and cardiovascular disease,[24] compared with their unexposed, same-sex siblings. Other studies show that monozygotic twins with discordant birth weight show whole-blood DNA methylation differences.[25] In rural Gambia, where there are profound seasonal fluctuations in the availability of food, whole-blood DNA methylation patterns in children varied according to the season of their conception.[26] More research is needed to determine whether these methylation changes associated with impaired fetal growth mediate high cardiometabolic risk later in life. The identification of high-risk epigenetic signatures may point to metabolic pathways affected by fetal undernutrition. Upon successful validation, these epigenetic patterns may also serve as biomarkers for cardiovascular risk stratification.

Any discussion of cardiovascular epigenetics would be remiss to exclude miRNAs, reviewed in 2 recent articles published in *Revista Española de Cardiología*.[27,28] Several large-scale studies[29–31] have offered new insights into the role of miRNAs (miR148a and miR33a/b) as key regulators of metabolic processes, particularly fatty acid oxidation and cholesterol efflux; other investigations have proposed the use of miRNAs (eg, the miR133 family, miR19b-3p, miR134-5p, and miR-186-5p) as biomarkers of myocardial infarction.[32,33] These findings, however, have yet to be translated to the clinical setting, and most studies have so far failed to show the superiority of miRNA-based biomarkers over traditional risk factors such as troponin.[34] Additionally, miRNAs exhibit impressively pleiotropic effects, impacting expression of multiple genes in a variety of tissues, thus posing the challenge of specificity for any potential therapies. Even more interestingly, miRNAs (specifically miR33a/b) play a role in interactions between 3 genes that have previously emerged on methylome-wide screens for plasma lipids (*CPT1A*, *ABCG1*, and *SREBF1*),[30] illustrating the necessity of analytic approaches that integrate across -omics layers for a fuller understanding of complex traits.

Another -omic layer that has received attention in the cardiovascular realm is metabolomics, the quantification of small-molecule circulating metabolites (usually via nuclear magnetic resonance or mass spectrometry) that offers additional granularity in investigating disease etiology. For example, 4 metabolites were predictive of cardiovascular risk in multiple cohorts, with serum phenylalanine and monounsaturated fatty acid levels indicating a higher likelihood of incident events, while omega-6 and docoxenaenoic acids were associated with an improved risk profile,[35] independently of traditional risk factors. While the physiologic relevance of these 4 markers had been relatively well understood even prior to that study, other high-throughput metabolomics screens have identified novel targets, most notably GlycA,[36] a marker of systemic inflammation that was subsequently linked to cardiovascular morbidity and mortality,[37] and trimethylamine-N-oxide (TMAO), a proatherogenic species[38] that promoted atherosclerosis in a mouse model[39] and was independently associated with the risk of adverse cardiovascular events[40] in humans. The TMAO discovery in particular illustrated the contribution of yet another -omic layer: metagenomics, or the composition of bacteria that live in the gut and synthesize TMAO precursors in response to dietary inputs such as red meat, fish, and eggs.[41] Current efforts are underway to develop drugs targeting the microbiome; although still in their infancy, they illustrate the clinical promise of well-validated -omic targets.

With the growing abundance of -omic data on hundreds of thousands of individuals, the main challenges facing cardiovascular genomics lie in analysis, interpretation, and application. In addition to the obvious hurdle of the astronomical multiple testing burden, there is a dearth of methods that fully exploit the wealth of quantified variation by integrating across -omics layers. Most cardiovascular studies that do attempt integration do it in a 'pairwise' fashion, eg, linking DNA sequence and epigenetics/expression via methylation/expression quantitative trait analysis, or via GWAS of metabolomics traits or microbiome composition, or via Mendelian randomization. These 'pairwise' methods comparing 2 -omic layers fail to capture the contribution of other intermediate phenotypes, as well as higher-order interactions. Structural equation models[42] represent a more sophisticated strategy for generating causal insights into multidimensional data, yet—much like Mendelian randomization—warrant their own set of assumptions that may not be explicitly tested using data from human populations. A number of integration methods that rely on available bioinformatics resources, eg, methods that leverage prior knowledge of biologic pathways, are limited by the inconsistency of public databases and their bias toward known genes.[43]

The metaphor of 'drinking from a fire hose' that emerged during early GWAS[44] is even more accurate today, and methodological creativity is urgently needed to harness the tremendous potential of the -omics data that are now or will soon become available. To transform cardiovascular -omics studies from a very expensive fishing expedition to truly personalized medicine, the field needs fully integrative trans-omic approaches, complete with laboratory follow-up of -omics findings. Mere replication of -omics results is no longer sufficient but must be supplemented by functional validation. If successful, such studies could inform novel drug therapies and risk stratification approaches, leveraging 'big data' to make a big impact in the fight against cardiovascular disease.

CONFLICTS OF INTEREST

None declared.

REFERENCES

1. National Heart, Lung, and Blood Institute. Trans-omics for Precision Medicine (TOPMed Program): Whole Genome Sequencing (WGS) Project [accessed 8 Mar 2017]. Available at: https://www.nhlbi.nih.gov/research/resources/nhlbi-precision-medicine-initiative/topmed/wgs.
2. National Human Genome Research Institute. NHGRI Genome Sequencing Program (GSP): Centers for Common Disease Genomics [accessed 8 Mar 2017]. Available at: https://www.genome.gov/27563570/.
3. Keshavan M [STAT]. Illumina says it can deliver a $100 genome – soon [accessed 8 Mar 2017]. Available at: https://www.statnews.com/2017/01/09/illumina-ushering-in-the-100-genome/.

4. Teslovich TM, Musunuru K, Smith AV, et al. Biological, clinical and population relevance of 95 loci for blood lipids. *Nature.* 2010;466:707–713.
5. Global Lipids Genetics C, Willer CJ, Schmidt EM, et al. Discovery and refinement of loci associated with lipid levels. *Nat Genet.* 2013;45:1274–1283.
6. Wain LV, Verwoert GC, O'Reilly PF, et al. Genome-wide association study identifies six new loci influencing pulse pressure and mean arterial pressure. *Nat Genet.* 2011;43:1005–1011.
7. Johnson AD, Newton-Cheh C, Chasman DI, et al. Association of hypertension drug target genes with blood pressure and hypertension in 86,588 individuals. *Hypertension.* 2011;57:903–910.
8. Salfati E, Morrison AC, Boerwinkle E, Chakravarti A. Direct Estimates of the Genomic Contributions to Blood Pressure Heritability within a Population-Based Cohort (ARIC). *PLoS One.* 2015;10:e0133031.
9. Nikpay M, Goel A, Won HH, et al. A comprehensive 1,000 Genomes-based genome-wide association meta-analysis of coronary artery disease. *Nat Genet.* 2015;47:1121–1130.
10. Kathiresan S, Srivastava D. Genetics of human cardiovascular disease. *Cell.* 2012;148:1242–1257.
11. Elosua R, Sayols-Baixeras S. The Genetics of Ischemic Heart Disease: From Current Knowledge to Clinical Implications. *Rev Esp Cardiol.* 2017;70:754–762.
12. Talmud PJ, Smart M, Presswood E, et al. ANGPTL4 E40K and T266 M: effects on plasma triglyceride and HDL levels, postprandial responses, and CHD risk. *Arterioscler Thromb Vasc Biol.* 2008;28:2319–2325.
13. TG, HDL Working Group of the Exome Sequencing Project NHLBlood I, Crosby J, et al. Loss-of-function mutations in APOC3, triglycerides, and coronary disease. *N Engl J Med.* 2014;371:22–31.
14. Heard E, Martienssen RA. Transgenerational epigenetic inheritance: myths and mechanisms. *Cell.* 2014;157:95–109.
15. Sen A, Heredia N, Senut MC, et al. Multigenerational epigenetic inheritance in humans: DNA methylation changes associated with maternal exposure to lead can be transmitted to the grandchildren. *Sci Rep.* 2015;5:14466.
16. Irvin MR, Zhi D, Joehanes R, et al. Epigenome-wide association study of fasting blood lipids in the Genetics of Lipid-lowering Drugs and Diet Network study. *Circulation.* 2014;130:565–572.
17. Braun KV, Dhana K, de Vries PS, et al. Epigenome-wide association study (EWAS) on lipids: the Rotterdam Study. *Clin Epigenetics.* 2017;9:15.
18. Sayols-Baixeras S, Subirana I, Lluis-Ganella C, et al. Identification and validation of seven new loci showing differential DNA methylation related to serum lipid profile: an epigenome-wide approach. The REGICOR study. *Hum Mol Genet.* 2016;25:4556–4565.
19. Aslibekyan S, Agha G, Ligthart S, et al. Novel DNA Methylation Loci Associated With Circulating Tumor Necrosis Factor-alpha, a Marker of Systemic Inflammation. *Circulation.* 2016;134:A18708.
20. Kraft P. Curses–winner's and otherwise–in genetic epidemiology. *Epidemiology.* 2008;19:649–651.
21. Dekkers KF, van Iterson M, Slieker RC, et al. Blood lipids influence DNA methylation in circulating cells. *Genome Biol.* 2016;17:138.
22. Low FM, Gluckman PD, Hanson MA. Developmental plasticity and epigenetic mechanisms underpinning metabolic and cardiovascular diseases. *Epigenomics.* 2011;3:279–294.
23. Heijmans BT, Tobi EW, Stein AD, et al. Persistent epigenetic differences associated with prenatal exposure to famine in humans. *Proc Natl Acad Sci U S A.* 2008;105:17046–17049.
24. Tobi EW, Lumey LH, Talens RP, et al. DNA methylation differences after exposure to prenatal famine are common and timing- and sex-specific. *Hum Mol Genet.* 2009;18:4046–4053.
25. Chen M, Baumbach J, Vandin F, et al. Differentially Methylated Genomic Regions in Birth-Weight Discordant Twin Pairs. *Ann Hum Genet.* 2016;80:81–87.
26. Waterland RA, Kellermayer R, Laritsky E, et al. Season of conception in rural gambia affects DNA methylation at putative human metastable epialleles. *PLoS Genet.* 2010;6:e1001252.
27. Corella D, Ordovás JM. Basic Concepts in Molecular Biology Related to Genetics and Epigenetics. *Rev Esp Cardiol.* 2017;70:744–753.
28. De Gonzalo-Calvo D, Iglesias-Gutiérrez E, Llorente-Cortés V. Epigenetic Biomarkers and Cardiovascular Disease: Circulating MicroRNA. *Rev Esp Cardiol.* 2017;70:763–769.
29. Goedeke L, Rotllan N, Canfran-Duque A, et al. MicroRNA-148a regulates LDL receptor and ABCA1 expression to control circulating lipoprotein levels. *Nat Med.* 2015;21:1280–1289.
30. Pfeiffer L, Wahl S, Pilling LC, et al. DNA methylation of lipid-related genes affects blood lipid levels. *Circ Cardiovasc Genet.* 2015;8:334–342.
31. Wagschal A, Najafi-Shoushtari SH, Wang L, et al. Genome-wide identification of microRNAs regulating cholesterol and triglyceride homeostasis. *Nat Med.* 2015;21:1290–1297.
32. Boon RA, Dimmeler S. MicroRNAs in myocardial infarction. *Nat Rev Cardiol.* 2015;12:135–142.
33. Wang KJ, Zhao X, Liu YZ, et al. Circulating MiR-19b-3p, MiR-134-5p and MiR-186-5p are Promising Novel Biomarkers for Early Diagnosis of Acute Myocardial Infarction. *Cell Physiol Biochem.* 2016;38:1015–1029.
34. Goretti E, Devaux Y. Which future for circulating microRNAs as biomarkers of acute myocardial infarction? *Ann Transl Med.* 2016;4:440.
35. Wurtz P, Havulinna AS, Soininen P, et al. Metabolite profiling and cardiovascular event risk: a prospective study of 3 population-based cohorts. *Circulation.* 2015;131:774–785.
36. Otvos JD, Shalaurova I, Wolak-Dinsmore J, et al. GlycA: A Composite Nuclear Magnetic Resonance Biomarker of Systemic Inflammation. *Clin Chem.* 2015;61:714–723.
37. Duprez DA, Otvos J, Sanchez OA, Mackey RH, Tracy R, Jacobs Jr DR. Comparison of the Predictive Value of GlycA and Other Biomarkers of Inflammation for Total Death, Incident Cardiovascular Events, Noncardiovascular and Noncancer Inflammatory-Related Events, and Total Cancer Events. *Clin Chem.* 2016;62:1020–1031.
38. Mayr M. Recent highlights of metabolomics in cardiovascular research. *Circ Cardiovasc Genet.* 2011;4:463–464.
39. Wang Z, Klipfell E, Bennett BJ, et al. Gut flora metabolism of phosphatidylcholine promotes cardiovascular disease. *Nature.* 2011;472:57–63.
40. Tang WH, Wang Z, Levison BS, et al. Intestinal microbial metabolism of phosphatidylcholine and cardiovascular risk. *N Engl J Med.* 2013;368:1575–1584.
41. Koeth RA, Wang Z, Levison BS, et al. Intestinal microbiota metabolism of L-carnitine, a nutrient in red meat, promotes atherosclerosis. *Nat Med.* 2013;19:576–585.
42. Tseng GC, Debashis G, Zhou XJ. *Structure Equation Models In Integrating Omics Data.* Cambridge: Cambridge University Press; 2015:285–288.
43. Aslibekyan S, Almeida M, Tintle N. Pathway analysis approaches for rare and common variants: insights from Genetic Analysis Workshop 18. *Genet Epidemiol.* 2014;38(Suppl 1):S86–S91.
44. Hunter DJ, Kraft P. Drinking from the fire hose–statistical issues in genomewide association studies. *N Engl J Med.* 2007;357:436–439.