Published in final edited form as:

Rev Esp Cardiol (Engl Ed). 2017 September ; 70(9): 696–698. doi:10.1016/j.rec.2017.03.012.

The State of Cardiovascular Genomics: Abundant Data, Limited Information

Stella ASLIBEKYAN^{a,*} and Edward A. RUIZ-NARVÁEZ^b

^aDepartment of Epidemiology, University of Alabama at Birmingham, Birmingham, Alabama, United States

^bSlone Epidemiology Center, Boston University, Boston, Massachusetts, United States

“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 Medicine¹ and the Centers for Common Disease Genomics² programs in the United States are poised to sequence more than ~250,000 individual genomes, of which ~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 horizon³.

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%-12% of the overall trait variance and 25%-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 ~2% of the phenotypic variance.⁸ The largest meta-analysis of coronary artery disease GWAS to date revealed a similar genetic

*Corresponding author: Department of Epidemiology, University of Alabama at Birmingham, 1665 University Blvd, RPHB 230J, Birmingham, AL 35205, United States. stella7@gmail.com (S. Aslibekyan).

CONFLICTS OF INTEREST

None declared.

architecture, with most of the heritability (~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 GWAS⁴ 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 two review papers published in *Revista Española de Cardiología*, is disproportionate to the phenotypic variance explained (< 1% in plasma triglyceride levels).¹¹ 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*¹²—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,^{13,14} it potentially may 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%-11.6% of triglyceride variation across several cohorts;¹⁵⁻¹⁷ several consortia meta-analyses of other cardiovascular phenotypes (incident coronary heart disease, circulating cytokines, hypertension, and others) are currently underway and have shown preliminary promising results.¹⁸ Although this relative success of the epigenome-wide studies may in part be due to the 'winner's curse,'¹⁹ it also has biological underpinnings: methylation variation, with its corresponding changes in gene transcription, is more proximal to the phenotype 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 (e.g. liver for lipid metabolism). Recent attempts to address the causality challenge using Mendelian randomization techniques²⁰ were limited by the lack of robust genetic instruments for complex traits like plasma lipids or blood pressure, hampered by the low proportion of variance explained by known polymorphisms as discussed above. Therefore, there is pressing demand for carefully controlled functional studies *in vitro* and 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.^{20,21} A consistent body of evidence shows that impaired fetal growth is associated with DNA methylation levels. For example, subjects who suffered 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 six decades later,²² as well as in 6 out of 15 other candidate genes involved in metabolic and cardiovascular disease,²³ compared with their unexposed, same-sex siblings. Other studies

show that monozygotic twins with discordant birth weight show whole-blood DNA methylation differences.²⁴ 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.²⁵ 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 a recent article published in *Revista Española de Cardiología*. Several large-scale studies²⁶⁻²⁸ 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 (e.g. the miR133 family, miR19b-3p, miR134-5p, and miR-186-5p) as biomarkers of myocardial infarction.^{29, 30} These findings, however, have yet to be translated to the clinical setting, and the majority of studies have so far failed to show superiority of miRNA-based biomarkers over traditional risk factors such as troponin.³¹ Additionally, miRNAs exhibit impressively pleiotropic effects, impacting expression of multiple genes in a variety of tissues, posing the challenge of specificity for any potential therapies. Even more interestingly, miRNAs (specifically miR33a/b) play a role in interactions between three genes that have previously emerged on methylome-wide screens for plasma lipids (*CPT1A*, *ABCG1*, and *SREBF1*),²⁷ 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, four metabolites were predictive of cardiovascular risk in multiple cohorts, with serum phenylalanine and monounsaturated fatty acid levels indicating higher likelihood of incident events, while omega-6 and docoxenaenoic acids were associated with an improved risk profile,³² independently of traditional risk factors. While the physiologic relevance of these four markers had been relatively well understood even prior to that study, other high-throughput metabolomics screens have identified novel targets, most notably GlycA,³³ a marker of systemic inflammation that was subsequently linked to cardiovascular morbidity and mortality,³⁴ and trimethylamine-N-oxide (TMAO), a pro-atherogenic species³⁵ that promoted atherosclerosis in a mouse model³⁶ and was independently associated with the risk of adverse cardiovascular events³⁷ 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.³⁸ 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, e.g. 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 two –omic layers fail to capture the contribution of other intermediate phenotypes, as well as higher-order interactions. Structural equation models³⁹ 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, e.g. methods that leverage prior knowledge of biologic pathways, are limited by the inconsistency of public databases and their bias towards known genes.⁴⁰

The metaphor of ‘drinking from a fire hose’ that emerged during early GWAS⁴¹ 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.

References

1. National Heart Lung Blood Institute. Trans-omics for Precision Medicine: Whole Genome Sequencing Project. Jul 18. 2016 Retrieved March 8, 2017, from <https://www.nhlbi.nih.gov/research/resources/nhlbi-precision-medicine-initiative/topmed/wgs>.
2. National Human Genome Research Institute. NHGRI Genome Sequencing Program, Centers for Common Disease Genomics. Mar 15. 2016 Retrieved March 8, 2017, from <https://www.genome.gov/27563570/>.
3. Keshavan, M. Illumina says it can deliver a \$100 genome - soon. STAT. Jan 9. 2017 Retrieved March, 8, 2017, from <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. [PubMed: 20686565]
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. [PubMed: 24097068]
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. [PubMed: 21909110]
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. [PubMed: 21444836]
8. Salafati 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. [PubMed: 26162070]

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. [PubMed: 26343387]
10. Kathiresan S, Srivastava D. Genetics of human cardiovascular disease. *Cell.* 2012; 148:1242–1257. [PubMed: 22424232]
11. Talmud PJ, Smart M, Presswood E, et al. ANGPTL4 E40K and T266M: effects on plasma triglyceride and HDL levels, postprandial responses, and CHD risk. *Arterioscler Thromb Vasc Biol.* 2008; 28:2319–2325. [PubMed: 18974381]
12. TG, HDL Working Group of the Exome Sequencing Project NHL. Blood I, Crosby J, et al. Loss-of-function mutations in APOC3, triglycerides, and coronary disease. *N Engl J Med.* 2014; 371:22–31. [PubMed: 24941081]
13. Heard E, Martienssen RA. Transgenerational epigenetic inheritance: myths and mechanisms. *Cell.* 2014; 157:95–109. [PubMed: 24679529]
14. 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. [PubMed: 26417717]
15. 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; 130:565–72. [PubMed: 24920721]
16. Braun KV, Dhana K, de Vries PS, et al. Epigenome-wide association study (EWAS) on lipids: the Rotterdam Study. *Clin Epigenetics.* 2017; 9:15. [PubMed: 28194238]
17. Sayols-Baixeras S, Subirana I, Lluís-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. [PubMed: 28173150]
18. Aslibekyan S, Agha G, Ligthart S, et al. Novel DNA Methylation Loci Associated With Circulating Tumor Necrosis Factor- α , a Marker of Systemic Inflammation. *Circulation.* 2016; 134:A18708.
19. Kraft P. Curses–winner’s and otherwise–in genetic epidemiology. *Epidemiology.* 2008; 19:649–651. [PubMed: 18703928]
20. Dekkers KF, van Iterson M, Sliker RC, et al. Blood lipids influence DNA methylation in circulating cells. *Genome Biol.* 2016; 17:138. [PubMed: 27350042]
21. Low FM, Gluckman PD, Hanson MA. Developmental plasticity and epigenetic mechanisms underpinning metabolic and cardiovascular diseases. *Epigenomics.* 2011; 3:279–294. [PubMed: 22122338]
22. 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. [PubMed: 18955703]
23. 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. [PubMed: 19656776]
24. 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. [PubMed: 26831219]
25. 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. [PubMed: 21203497]
26. 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. [PubMed: 26437365]
27. 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. [PubMed: 25583993]
28. 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. [PubMed: 26501192]

29. Boon RA, Dimmeler S. MicroRNAs in myocardial infarction. *Nat Rev Cardiol*. 2015; 12:135–142. [PubMed: 25511085]
30. 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. [PubMed: 26939053]
31. Goretti E, Devaux Y. Which future for circulating microRNAs as biomarkers of acute myocardial infarction? *Ann Transl Med*. 2016; 4:440. [PubMed: 27942531]
32. 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. [PubMed: 25573147]
33. 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. [PubMed: 25779987]
34. Duprez DA, Otvos J, Sanchez OA, Mackey RH, Tracy R, Jacobs DR Jr. 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; 62:1020–31. [PubMed: 27173011]
35. Mayr M. Recent highlights of metabolomics in cardiovascular research. *Circ Cardiovasc Genet*. 2011; 4:463–464. [PubMed: 21846870]
36. Wang Z, Klipfell E, Bennett BJ, et al. Gut flora metabolism of phosphatidylcholine promotes cardiovascular disease. *Nature*. 2011; 472:57–63. [PubMed: 21475195]
37. Tang WH, Wang Z, Levison BS, et al. Intestinal microbial metabolism of phosphatidylcholine and cardiovascular risk. *N Engl J Med*. 2013; 368:1575–1584. [PubMed: 23614584]
38. 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. [PubMed: 23563705]
39. Tseng, GC., Debashis, G., Zhou, XJ. Integrating Omics Data. Cambridge: Cambridge University Press; 2015. Structure Equation Models; p. 285-288.
40. 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):S86–91. [PubMed: 25112195]
41. Hunter DJ, Kraft P. Drinking from the fire hose—statistical issues in genomewide association studies. *N Engl J Med*. 2007; 357:436–439. [PubMed: 17634446]