Circulation

EDITORIAL

Linking Genetics and Proteomics: Gene-Protein Associations Built on Diversity

Heribert Schunkert, MD; Manuel Mayr, MD, PhD

Medical applications from big data science emerge predominantly from the discovery of associations, for example, between a genomic variant and a specific disease. Although the starting point of such studies is merely observational, we have been witnessing unprecedented advances in big data science during the last 15 years, which also allow mechanistic inferences. Indeed, identification of hundreds of genetic variants showing disease associations helped better map the complex genetic architecture of many cardiovascular diseases. On the basis of such big data, Mendelian randomization studies can inform which risk markers are causal risk factors, whereas polygenic risk scores can be used to predict disease risk. Genetic associations, however, will benefit from other layers of molecular information, in particular with regard to prioritizing variants for further functional investigations. Transcriptomics, proteomics, and metabolomics will be essential to explore dysregulated pathways that ultimately cause disease. Such integrated big data sets may also have commercial value, ie, decisions on drug developments can be guided by linking potential targets to causal pathways depicted from gene variants to dysfunctional proteins. Indeed, the promise of precision medicine builds on such insights leading to differentiation of patient subgroups that benefit from different therapeutic strategies.[1]

Article, see p 357

Big data science has been enabled by high-throughput methods for data generation and novel computational approaches to analyze large datasets. Genomics has been at the forefront of these developments and now expands into whole exome or even whole genome sequencing in large cohorts. Transcriptomics is moving toward single-cell analysis. Proteomics and metabolomics are about to complete the information on the key molecular layers of biological systems at the large scale.

In terms of the cohorts for big data science, an in-depth and accurate phenotypic assessment along with long-term follow-up have proven to be most valuable. And, of course, size matters. Not withstanding the importance of studying rare diseases, the principle "bigger is better" holds true for improving the resolution for the identification of disease associations. Only at scale challenges in big data science, such as adjustment for multiple testing and confounding, can be addressed successfully.

However, human diversity provides another opportunity to search for associations between genomic markers and phenotypic variability.[2] In fact, genetic epidemiological research has long focused on rather homogeneous cohorts in which numerous genes may lack functionally relevant variants. Given the stringent statistical thresholds in big data science, functionally relevant associations might have been overlooked. Thus, to take advantage of the gene pool of the global population—rather than predominantly of individuals of White ancestry—may further improve big data science in biomedical research (Figure).

The cradle of human evolution is the African continent, between 200 000 and 300 000 years ago. Starting about 70 000 to 100 000 years ago, the first modern humans migrated first northbound and subsequently spread around the globe, carrying with them a subset of the African gene pool. During these periods of migration and the associated environmental challenges, some genetic variation was lost. Therefore, African populations have been proposed to be particularly informative for

Key Words: Editorials ■ cardiovascular disease ■ ethnic diversity ■ genome-wide association study ■ genomics ■ proteomics

Correspondence to: Heribert Schunkert, MD, German Heart Center Munich, Technical University Munich, Lazarettstr. 36, 80636 Munich, Germany. Email schunkert@dhm.mhn.de

For Sources of Funding and Disclosures, see page 373.

Circulation is available at www.ahajournals.org/journal/circ

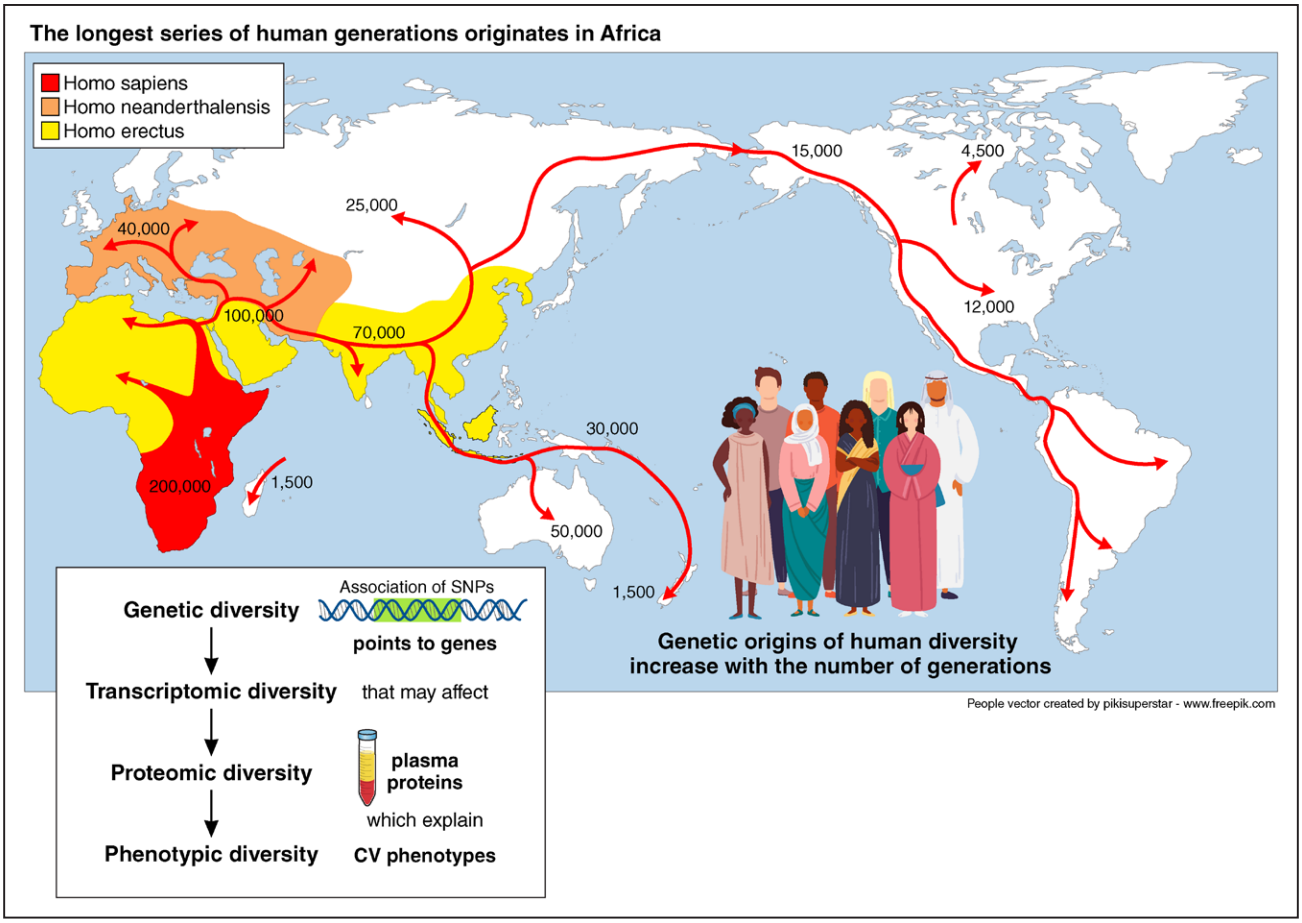

Figure. Genomic diversity improves resolution of association studies.
Starting in Africa, human evolution created—during many generations—the largest genetic diversity on this continent. Groups of people, taking some of the African gene pool with them, subsequently populated the globe and, to a small extent, interbred with other early human populations. Leveraging the full diversity of the human pool will help associate genetic variants with the full spectrum of transcriptomic, proteomic, and ultimately phenotypic variation. Katz et al[7] in this issue of *Circulation* filled a scientific gap in that they associated whole genome sequence with plasma proteomes from the Black population, revealing that about a quarter of significant associations were missed so far. CV indicates cardiovascular; and SNP, single nucleotide polymorphism.

investigation of human variation that arose among the early hominid predecessors between 100 000 and 5 million years ago.[2] It was hypothesized that the number of ancient variants and the selection pressures they survived may yield particularly rewarding insights into complex traits in all populations.

Other genomic variation was subsequently added, eg, the DNA of Neanderthals was only incorporated after modern humans left Africa.[3] Traces of Neanderthal DNA found in a human genome suggest that the first modern humans were interbreeding with our now-extinct relatives during migration. As a consequence, the full genetic and ultimately phenotypic diversity of the human population needs assessment of all ethnic groups. For example, the strongest genetic risk for coronary artery disease resides at the 9p21 locus—but is predominately found in Western Europeans.[4] In fact, this risk haplotype is largely absent in the Black population. [5] On the other hand, common variants in APOL1 (apolipoprotein-L1), which protect against sleeping sickness but increase risk of renal failure, are predominately found in sub-Saharan populations.[6]

An article in this issue of *Circulation* contributes to mapping the Black map of genome-proteome associations.[7] Katz and colleagues integrated whole genome DNA sequencing and proteomic profiling of ≈1300 proteins in the plasma of individuals of Black descent. They report 569 associations between genomic variants and plasma proteins that reach a Bonferroni-adjusted significance level.[7] One in 3 proteins found in the plasma was quantitatively modulated by genetic variants, of which about two-thirds were in *cis*, ie, in the vicinity of respective genes, and one-third was in *trans*.[7] Overall, about a third of the variability of plasma protein concentrations was found the be partially heritable on the basis of the 28 million genomic variants studied. In this respect, Sun et al had previously reported a far lower heritability (8%).[8]

It is important to note that the authors validated the proteomic profiling by the SOMAScan expanded platform in 2 other cohorts, in which ≈90% of hits were consistent

in directionality. They also validated the profiling using another proteomic platform (Olink Explore), in which 86% of hits were in similar direction, with 51% of all associations confirmed at a Bonferroni-adjusted *P* value. Both the SOMAScan and the Olink platform rely on binders for relative protein quantitation: the SOMAScan assay uses aptamers as a single binder. The Olink Explore platform uses dual antibody-based proximity extension assays. For studying the effects of genomic variants on protein levels, it is essential to exclude that the genomic variants induce platform-specific binding effects as illustrated by some highly discordant results. By replicating parts of the results of the SOMAScan using the Olink platform, the concordant results can be considered more reliable. However, even concordant results between 2 platforms may be misleading if aptamer and antibody binding occur in the same protein region that is affected by the genetic variant. This possibility should not be readily discounted and could be more reliably addressed by proteomics profiling techniques that do not rely on binders. Mass spectrometry measures peptides directly, and assays can be designed for the variant and the wild-type protein using authentic reference peptides for accurate quantitation.[9]

It is important to note that a quarter of associations were novel and missed by investigations in other ethnic groups. Several of the newly identified associations represent biologically interesting candidates.

From a functional perspective, the findings include a novel association linking the *APOE* gene locus with *ZAP70* (zeta chain of T cell receptor–associated protein kinase 70) and *MMP-3* (matrix metalloproteinase-3) protein levels in plasma, as well as a novel pleiotropic locus at the *HPX* (hemopexin) gene, which is associated with 9 proteins. Hemopexin may be of particular interest in Black individuals with sickle-cell disease because this gene has been found to have heme-scavenging properties leading to reduced inflammation in mouse models of sickle-cell disease.[10]

Of particular clinical interest are potential novel insights on the genetic susceptibility to chronic kidney disease and cardiac amyloid deposition through variants at the *APOL1* and *ATTR* loci that revealed African-specific associations with plasma proteins. Amyloidosis may result from the V122I mutation in the TTR gene leading to misfolding of the transthyretin tetramer, ultimately resulting in abnormal protein deposition in myocardium and nerve tissue. This variant, occurring in 3% to 4% of Black individuals, was found to be a robust protein quantitative trait locus for RBP4 (retinol-binding protein 4), a binding partner of TTR, a finding that may allude to novel mechanisms involved in the pathogenesis of cardiomyopathy or neuropathy.

A variant at the *APOL1* locus, rs73885319, was found to have a minor allele frequency of 23% in the Black population, whereas the variant is not present in individuals of European ancestry. In addition to being associated with levels of APOL1, Katz et al[7] observed the sentinel single nucleotide polymorphism at this locus also determines plasma levels of CKAP2 (cytoskeleton-associated protein 2), which has been linked to tumor formation and renal tubular necrosis.[6]

All these associations between genomic variants and plasma proteins in people of African descent can only be a starting point for further mechanistic explorations. In this respect, it will be important to make these findings accessible for the broader research community, as outlined by the authors.[7] Moreover, functional studies should follow to corroborate the findings and characterize the pathways that link genetic variations and protein levels. Also, it has to be taken into consideration that genetic variants tend to explain only a small proportion of the dynamic range of protein concentrations in plasma. Other factors, such as plasma protein synthesis, proteolysis, and renal clearance, are stronger determinants of plasma protein abundance and may lead to spurious associations.[11] Thus, follow-up experiments are needed to address biological plausibility, pleiotropic effects, and possible pathways to disease to enable clinical translation.

ARTICLE INFORMATION

Affiliations

Department of Cardiology, Deutsches Herzzentrum München, Technische Universität München, Munich, Germany (H.S.). Deutsches Zentrum für Herz- und Kreislauferkrankungen, Partner Site Munich Heart Alliance, Munich, Germany (H.S.). King's College London British Heart Foundation Centre, School of Cardiovascular Medicine and Sciences, United Kingdom (M.M.).

Sources of Funding

H.S. and M.M. acknowledge support from the Foundation Leducq (PlaqueOmics: Novel Roles of Smooth Muscle and Other Matrix Producing Cells in Atherosclerotic Plaque Stability and Rupture, 18CVD02). H.S. received grants from the European Union under grant agreement HEALTH-F2-2013-601456 (CVgenes-at-target), the British Heart Foundation–German Center of Cardiovascular Research collaboration, and the ERA-NET (Druggable-MI-Genes: 01KL1802), and was also supported by grants from the Federal German Ministries (AbCD-Net, grant 01ZX1706C; BLOCK-CAD, grant 16GW0198K; ModulMax, grant ZF4590201BA8), and the DFG as part of the Sonderforschungsbereich CRC 1123 (B2), the Transregio TRR 267 (B05), as well as the DigiMed Bayern project (DBM-1805-0001). M.M. is a British Heart Foundation (BHF) Chair Holder (CH/16/3/32406) with BHF program grant support (RG/16/14/32397). M.M. was awarded a BHF Special Project grant to participate in the ERA-CVD Transnational Grant "MacroERA: Noncoding RNAs in Cardiac Macrophages and Their Role in Heart Failure" and is part of the Marie Skłodowska-Curie Innovative Training Network TRAIN- HEART (http://train-heart.eu) as well as the excellence initiative VASCage (Center for Promoting Vascular Health in the Ageing Community, project number 868624) of the Austrian Research Promotion Agency FFG (COMET program–Competence Centers for Excellent Technologies) funded by the Austrian Ministry for Transport, Innovation and Technology; the Austrian Ministry for Digital and Economic Affairs; and the federal states Tyrol (via Standortagentur), Salzburg, and Vienna (via Vienna Business Agency). M.M.'s work was supported by the National Institute of Health Research Biomedical Research Center based at Guy's and St Thomas' National Health Service Foundation Trust and King's College London in partnership with King's College Hospital.

Disclosures

None.

REFERENCES

1. Chen Z, Schunkert H. Genetics of coronary artery disease in the post-GWAS era. *J Intern Med*. 2021;290:980–992. doi: 10.1111/joim.13362

2. McClellan JM, Lehner T, King MC. Gene discovery for complex traits: lessons from Africa. *Cell*. 2017;171:261–264. doi: 10.1016/j.cell.2017.09.037
3. Higham T, Douka K, Wood R, Ramsey CB, Brock F, Basell L, Camps M, Arrizabalaga A, Baena J, Barroso-Ruíz C, et al. The timing and spatiotemporal patterning of Neanderthal disappearance. *Nature*. 2014;512:306–309. doi: 10.1038/nature13621
4. Samani NJ, Erdmann J, Hall AS, Hengstenberg C, Mangino M, Mayer B, Dixon RJ, Meitinger T, Braund P, Wichmann HE, et al. Genomewide association analysis of coronary artery disease. *N Engl J Med*. 2007;357:443–453. doi: 10.1056/NEJMoa072366
5. Tcheandjieu C, Assimes T, Zhu X, Hilliard A, Clarke C, Napolioni V, Ma S, Fang H, Gorman BR, Min Lee K, et al. A large-scale multi-ethnic genome-wide association study of coronary artery disease. Accessed December 11, 2021. https://assets.researchsquare.com/files/rs-275591/v1_covered.pdf?c=1631857985.
6. Genovese G, Friedman DJ, Ross MD, Lecordier L, Uzureau P, Freedman BI, Bowden DW, Langefeld CD, Oleksyk TK, Uscinski Knob AL, et al. Association of trypanolytic ApoL1 variants with kidney disease in African Americans. *Science*. 2010;329:841–845. doi: 10.1126/science.1193032
7. Katz DH, Tahir UA, Bick AG, Pampana A, Ngo D, Benson MD, Yu Z, Robbins JM, Chen Z-Z, Cruz DE, et al. Whole genome sequence analysis of the plasma proteome in Black adults provides novel insights into cardiovascular disease. *Circulation*. 2022;145:357–370. doi: 10.1161/CIRCULATIONAHA.121.055117
8. Sun BB, Maranville JC, Peters JE, Stacey D, Staley JR, Blackshaw J, Burgess S, Jiang T, Paige E, Surendran P, et al. Genomic atlas of the human plasma proteome. *Nature*. 2018;558:73–79. doi: 10.1038/s41586-018-0175-2
9. Joshi A, Mayr M. In aptamers they trust: the caveats of the SOMAscan Biomarker discovery platform from SomaLogic. *Circulation*. 2018;138:2482–2485. doi: 10.1161/CIRCULATIONAHA.118.036823
10. Vinchi F, Costa da Silva M, Ingoglia G, Petrillo S, Brinkman N, Zuercher A, Cerwenka A, Tolosano E, Muckenthaler MU. Hemopexin therapy reverts heme-induced proinflammatory phenotypic switching of macrophages in a mouse model of sickle cell disease. *Blood*. 2016;127:473–486. doi: 10.1182/blood-2015-08-663245
11. Joshi A, Rienks M, Theofilatos K, Mayr M. Systems biology in cardiovascular disease: a multiomics approach. *Nat Rev Cardiol*. 2021;18:313–330. doi: 10.1038/s41569-020-00477-1

Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China

Union Hospital, Tongji Medical College, Huazhong University of Science and Technology was founded on September 8, 1866, by Griffith John (**top left**) from the UK. Originally named Hankow Renji Hospital with the meaning of "healing the world with love and uniting the people with sincerity", the hospital was located at Houhualouju Alley (now Jiaotong Alley) and was equipped originally with 2 wards, 15 beds, and 1 part-time doctor (**top middle**). In 1891, a branch facility named Margaret Hospital was established for female patients (**top right**). In 1902, Thomas Gillison, President of Renji Hospital, and physician Percy Lonsdale McAll co-founded Union Medical College, which was the first medical school of higher education established in Hankow. In 1928, Boyi Health Technical College founded by George Hadden, a British doctor, was transferred into Hankow Union Hospital to set up its headquarters (**bottom left**).

In May 2000, the Hospital was renamed Union Hospital (**bottom middle**). As a major comprehensive facility (Main Campus, Cancer Center, West Campus, and Jinyinhu Campus), it hosts a total of 6000 beds with 6 700 000 outpatient visits, approximately 300 000 admissions, and 120 000 inpatient operations annually. Working under the jurisdiction of the National Commission of Health, Union Hospital, the oldest in Wuhan, integrates the functions of a medical service (a leading position in China in heart transplantation, heart-lung transplantation, conjoined twin separation, bone marrow transplantation, laparoscopic colectomy for Hirschsprung's disease) and scientific research (7751 staff members, 10 national key disciplines, and 25 national key clinical specialties). Moreover, Union Hospital is the first center in the world to fight the COVID-19 pandemic. In 2021, Union Hospital will celebrate its 155th anniversary commemorated with a national stamp (**bottom right**).

Photographs courtesy of the Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China.

Pin Ye, MD
Chuanqi Cai, MD, PhD
Department of Vascular Surgery, Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China