Translational Epidemiology
Entering a Brave New World of Team Science

Ravi Shah, Alexander R. Pico, Jane E. Freedman

The measure of greatness in a scientific idea is the extent to which it stimulates thought and opens up new lines of research.

—Paul A.M. Dirac

Translational epidemiology is the study of risk factors and diseases in large populations, harnessing the power of modern phenotyping, including large-scale omics approaches. In this review, we comment on the power and limitations of modern molecular translational epidemiology and suggest collaborative team science as a specific avenue to secure its existence in the next era of genomic research.

In the past 15 years, the emergence of faster, cheaper high-throughput methods to quantify circulating biomolecules and genetic signatures has transformed the landscape of cardiovascular investigation. At the heart of this, Big Data revolution resides the longitudinal cohort study, a collection of individuals exquisitely phenotyped across various axes of cardiovascular health and prospectively followed for prognosis and disease development. Since the 1960s, cohort-based epidemiology has provided landmarks in cardiovascular medicine (eg, smoking as a risk factor for coronary heart disease). More recently, a marriage between human cardiovascular epidemiology and high-throughput omics research has birthed translational epidemiology, based on the premise that the comprehensive integration of genetic, epigenetic, transcriptional, proteomic, and metabolic signatures with phenotypes informs disease and uncovers risk and prognostic factors, personalizing cardiovascular care. Because the initiative to grow prospective cohort studies from the thousands (Framingham, Jackson Heart Study, and Multi-Ethnic Study of Atherosclerosis) to the hundreds of thousands (eg, UK Biobank and Million Veterans Project) expands, it is imperative to define the benefits and limitations of translational epidemiology to best use the ongoing and newer cohort studies to the benefit of our patients.

The opinions expressed in this article are not necessarily those of the editors or of the American Heart Association.

From the Department of Medicine, Massachusetts General Hospital, Harvard Medical School, Boston (R.S.); Department of Bioinformatics, Gladstone Institutes, University of California, San Francisco (A.R.P.); and Department of Medicine, University of Massachusetts, Worcester (J.E.F.).

Correspondence to Ravi Shah, MD, Massachusetts General Hospital, Harvard Medical School, 55 Fruit St, Boston, MA 02114. E-mail rsvshah@partners.org; or Jane E. Freedman, MD, University of Massachusetts, Medical School, Plantation St, Worcester, MA 01605. E-mail jane.freedman@umassmed.edu


Circulation Research is available at http://circres.ahajournals.org

Genomics: The First Frontier of Translational Epidemiology

Early work establishing a link between genetic variation and phenotype relied on defined, extreme phenotypes (eg, familial hypercholesterolemia and the low-density lipoprotein receptor; hypertension and the renin–angiotensin system; and Jervell syndrome and variants in long QT syndrome genes) and candidate genes within families with these phenotypes. The paradigm was one of the pure translational bench-to-bedside clinical investigations that identify a clinical phenotype, map and sequence the gene, identify the variation and its phenotypic significance, discern mechanism with detailed molecular investigation, and ultimately translate these findings back to a clinical setting. We have since lived in a renaissance of rapid technological innovation for genome sequencing, from shotgun sequencing in early human genome efforts to imputation techniques from arrays and most recently next-generation sequencing approaches that enable interrogation of all coding sequences across a genome in a matter of days. The paradigm shifted from phenotype–genetics mechanism to broad genetic interrogation of a wide array of variants first to establish an association with complex phenotypes. Genetic epidemiologists have harnessed the power of phenotypes in large cohorts with blood-based markers and have collaborated in consortia to conduct genome-wide association studies to identify the genetic variants associated with nearly every cardiovascular and metabolic phenotype (eg, diabetes mellitus, obesity; coronary artery disease; atrial fibrillation, and hypertension) and imaging markers of subclinical atherosclerosis.

This departure from the standard scientific method—hypothesis then experiment, as opposed to its opposite—has generated the greatest backlash to modern translational epidemiology. There have been limitations cited in translational epidemiology, including replication, a large array of variants without functional confirmation, inability to link genetic variants with causal genes (or RNAs), type I error with multiplicity of modeling, and small effect sizes as primary limitations in translating the genetic epidemiology to clinical risk prediction. Although an advent in approach (Mendelian randomization within cohort studies and functional pathway analysis) begins to ascribe mechanism to genetic variants, collaborative interactions with biologists to understand function (via gene editing, pluripotent stem cell, or in vitro disease models) and bioinformatics expertise (via systems biology integration of genetic and epigenetic markers) will be essential to realize the power of large cohort translational epidemiology in the next phase of cohort-based genomic research.

Translational Epidemiology and Functional Biomolecules: Metabolites, Proteins, and RNA

Apart from genetic associations, innovations in mass spectrometry and sequencing approaches have directed translational
epidemiology toward variation in the circulating metabolome, transcriptome, and most recently the proteome in search for functional biomarkers and novel mechanisms of cardiovascular disease. Seminal investigations in large US- and European-based cohorts have established circulating branched–chain amino acids as markers of cardiometabolic disease independent of obesity, in line with mechanistic observations in smaller human and animal studies of their role in insulin resistance in muscle tissue. In comparison, the prognostic importance of the plasma proteome has been reported in 2 major cohort studies so far, with preliminary data suggesting modest improvement in diagnostic sensitivity beyond traditional risk factors. In final, there is a growing interest in the role of the blood-based transcriptome in human disease. The importance of both protein- and nonprotein-coding genes in bridging genetics and cardiovascular risk is only beginning to be described.

In addition, the broad expression and relative stability of small circulating RNAs, including but not exclusive to microRNAs, suggests that these regulatory transcripts may have far-reaching importance in disease progression.

Each of these fields is too young to make conclusions as to its use to generate novel therapies or biomarkers of cardiovascular disease. Further validation and replication efforts across cohorts, specifically focusing on age, race, sex, and underlying disease burden, will be necessary to realize their full potential against known cardiovascular disease biomarkers. However, integrative translational epidemiological approaches to biomarker and pathway identification—uniting genetic, metabolic, and protein-based information—are central to fully harnessing the power of cohort-based investigation in understanding disease.

**Barriers and Solutions**

Certain barriers are inherent to translational epidemiology. First, does testing thousands of complex genetic, epigenetic, and metabolic markers of a disease state with defined phenotypes lead us to precision medicine or open us to false discovery? Collaborative efforts among bioinformatics expertise, statisticians, biologists, and epidemiologists in the field of systems biology have introduced methods for variable reduction, classification, and type I error control to begin to address some of these concerns. For example, the use of networks to define sets of interacting genes and proteins that can stratify patient outcomes, can serve as predictive biomarkers, and assist in interpretation of causal variants. Second, are any of these associations between circulating biomolecules or genetic variants functionally relevant in cardiovascular disease? Prioritizing candidates based on association between the different aspects of molecular physiology (eg, metabolite with genomic data) and on in vitro and in vivo models of disease will be critical in this regard. Third, what are the normative values of the biochemical profiles we are assaying (circadian variation, association with dietary pattern, sex, age, and race)? In certain, this has been a focus of investigation in several national consortia and will continue to be a deserving line of investigation. Fourth, are blood-based profiles enough? Is tissue of origin (or plasma compartment, eg, circulating exosomes) for each of these biomolecules relevant? Recent investigations have implicated circulating exosomes as containing a rich source of biomolecules mediating interaction between tissue types across the human organism, suggesting that tissue of origin and plasma compartment may be relevant to mechanism.

Our opinion is that an ideal study design harnesses the power of phenotyping within the large prospective cohorts to validate biomarkers uncovered during detailed investigations of well-defined phenotypes in smaller studies. These efforts should optimally be done in concert and linked at the point of funding. This approach begins in the strength of in-depth molecular characterization of a specific process (eg, acute heart failure, myocardial infarction, or sudden death) in a small cohort, followed by mechanistic biology to identify functionally important variants and ends in the use of detailed, quantitative coronary and vascular phenotypes and clinical end points in large prospective cohort studies. Furthermore, at a time when the significant limitations of mechanistic studies in model organisms are becoming apparent, it is important to acknowledge the fundamental and unique strengths of epidemiological cohorts: (1) diversity (eg, sex, race, and geographic based), (2) wealth of antecedent phenotypes (in some studies, 30 years of data on anthropometric, risk factor, and biochemical determinants of disease), (3) prospective event adjudication, and (4) surrogate, prognostic imaging end points (eg, coronary artery calcification and left ventricular mass). These features represent a unique investment over decades by participants, study coordinators, and governmental and foundation support and are not easily reproduced by prospective studies in similar numbers with similar follow-up without significant expenditure of time and cost.

Nevertheless, because the push to fund cohort studies via traditional granting mechanisms increases, it will be important to take a close look at what questions are important and whether we can fashion a chimeric data- and hypothesis-driven approach to translational epidemiology. In this context, several newer approaches that take advantage of longitudinal data and complex integrative phenotypes are garnering interest in the epidemiological community. Examining molecular predictors of lifetime risk exposures (eg, latent-class trajectory modeling) is a relevant avenue of investigation, specifically addressing how genetic and epigenetic markers can inform the evolution of cardiovascular risk, beginning in young adulthood. Applying systems biology approaches to phenotype data (eg, phenomapping) and patient similarity, networks are emerging areas of interest. In final, as US-based cohorts begin to age, investigations of relevant aging phenotypes with translation to animal models and small sample physiological studies may become an optimal use of the cohort study. These approaches will focus funding for translational epidemiology to questions of high impact for worldwide public health.

Ultimately, commitment from funding stakeholders and the scientific community by way of collaborative, multidisciplinary scientific teams will be the arbiter of survival or demise of the cohort study, and with it, translational epidemiology. This involves not only leveraging multidimensional data across different cohorts in a team science collaborative approach (eg, genomic, metabolomic, and epigenetic data) but more generally the wholehearted adoption of policies for data reuse and resource sharing. With the advent of newer biomarkers, newer high-throughput platforms, and bioinformatics to integrate the
available data for study in clinical trials and model systems, the

time is now to capitalize on decades of hard, collaborative work

in cohort-based investigation. We need to transform the investi-
gative paradigm in translational epidemiology to include closer

collaboration among basic, translational, and clinical investi-
gators to determine function and significance of the many as-
sociations we observe. Only then, as Professor Dirac suggests,

translational epidemiology will achieve greatness in stimulating

the next generation of thought in cardiovascular research.

Sources of Funding

The work was supported by the National Institutes of Health

K23HL127099 (to R. Shah) and National Institutes of Health UH3-

TR000921 and U01-HL126495 (to J.E. Freedman).

Disclosures

None.

References

1. Doyle JT, Dawber TR, Kannel WB, Kinch SH, Kahn HA. The re-

lationship of cigarette smoking to coronary heart disease: the sec-

ond report of the combined experience of the Albany, NY, and


jama.1964.03070230022006.

2. Speliotes EK, Willer CJ, Berndt SI, et al; MAGIC; Procardis Con-

sortium. Association analyses of 249,796 individuals reveal 18 new loci associated


3. Schunkert H, König IR, Kathiresan S, et al; Cardiogenics; CARDIoGRAM

Consortium. Large-scale association analysis identify-


2011;43:333–338. doi: 10.1038/ng.784.


357. doi: 10.1038/nature06007.

5. Ehret GB, Munroe PB, Rice KM, et al; International Consortium for

Blood Pressure Genome-Wide Association Studies; CARDIoGRAM

Consortium; CKDGen Consortium; KidneyGen Consortium; EchoGen

coronary; CHARGE-HF consortium. Genetic variants in novel path-


nm.2307.


Prevention Program Research Group. Metabotype profiles of diabetes in-

cidence and intervention response in the Diabetes Prevention Program.


ling DG, Williams SA. Development and validation of a protein-

based risk score for cardiovascular outcomes among patients with stable


jama.2016.5951.


ing reveals novel candidate biomarkers and pathways in cardio-


CIRCULATIONAHA.116.021803.

10. Freedman JE, Larson MG, Tanriverdi K, O’Donnell CJ, Morin K, Hakan-

son AS, Vasan RS, Johnson AD, Iafrati MD, Benjamin EJ. Relation of platelet and leukocyte inflammatory transcripts to body mass


doi: 10.1161/CIRCULATIONAHA.109.928192.


lular RNAs are widely detected in human plasma. Nat Commun.

2016;7:11106. doi: 10.1038/ncomms11106.

13. Shah SJ, Katz DH, Selvaraj S, Burke MA, Yancy CW, Gheorghiade

M, Bonow RO, Huang CC, Deo RC. Phenomapping for novel classi-

fication of heart failure with preserved ejection fraction. Circulation.


Key Words: genomics ■ metabolomics ■ research ■ risk factors ■ science
Translational Epidemiology: Entering a Brave New World of Team Science
Ravi Shah, Alexander R. Pico and Jane E. Freedman

Circ Res. 2016;119:1060-1062
doi: 10.1161/CIRCRESAHA.116.309881

The online version of this article, along with updated information and services, is located on the World Wide Web at:
http://circres.ahajournals.org/content/119/10/1060

Permissions: Requests for permissions to reproduce figures, tables, or portions of articles originally published in Circulation Research can be obtained via RightsLink, a service of the Copyright Clearance Center, not the Editorial Office. Once the online version of the published article for which permission is being requested is located, click Request Permissions in the middle column of the Web page under Services. Further information about this process is available in the Permissions and Rights Question and Answer document.

Reprints: Information about reprints can be found online at:
http://www.lww.com/reprints

Subscriptions: Information about subscribing to Circulation Research is online at:
http://circres.ahajournals.org/subscriptions/