The application of big data to cardiovascular disease: paths to precision medicine

Jane A. Leopold, …, Bradley A. Maron, Joseph Loscalzo

*J Clin Invest.* 2020;130(1):29-38. [https://doi.org/10.1172/JCI129203](https://doi.org/10.1172/JCI129203).

**Review Series**

Advanced phenotyping of cardiovascular diseases has evolved with the application of high-resolution omics screening to populations enrolled in large-scale observational and clinical trials. This strategy has revealed that considerable heterogeneity exists at the genotype, endophenotype, and clinical phenotype levels in cardiovascular diseases, a feature of the most common diseases that has not been elucidated by conventional reductionism. In this discussion, we address genomic context and (endo)phenotypic heterogeneity, and examine commonly encountered cardiovascular diseases to illustrate the genotypic underpinnings of (endo)phenotypic diversity. We highlight the existing challenges in cardiovascular disease genotyping and phenotyping that can be addressed by the integration of big data and interpreted using novel analytical methodologies (network analysis). Precision cardiovascular medicine will only be broadly applied to cardiovascular patients once this comprehensive data set is subjected to unique, integrative analytical strategies that accommodate molecular and clinical heterogeneity rather than ignore or reduce it.

Find the latest version:

[https://jci.me/129203/pdf](https://jci.me/129203/pdf)
The application of big data to cardiovascular disease: paths to precision medicine

Jane A. Leopold, Bradley A. Maron, and Joseph Loscalzo

Department of Medicine, Division of Cardiovascular Medicine, Brigham and Women’s Hospital and Harvard Medical School, Boston, Massachusetts, USA.

Introduction

Parallel advances and increased ease of access to high-throughput next-generation genomic, metabolomic, transcriptomic, and proteomic arrays coupled with widespread adaptation of the electronic health record heralded the era of big data in cardiovascular disease investigation. Rapid iteration of these technologies increasingly generates data outputs that are exponentially more complex compared with their earlier versions. This vast reservoir of clinical and biological information has revolutionized cardiovascular health and disease by establishing two contemporary needs that are germane to refining cardiovascular disease (end)phenotypes. First, there has been a shift in focus from data collection toward optimal and actionable data analysis and interpretation. Second, harnessing therapies for individualized genotypes and phenotypes, which collectively form the fundamental basis of precision medicine, has emerged as a defining challenge for big data (1).

Cardiovascular medicine stands to gain much by achieving these goals, particularly in the face of diminishing progress toward improving clinical outcomes in important, highly prevalent diseases such as heart failure (HF) with impaired diastolic relaxation, which is increasingly common in aging and obese populations (2–4). The evolving landscape of previously unrecognized comorbidities linked to cardiovascular diseases adds further to the challenge. Cardiovascular disease phenotypes, however, are complex entities that reflect a compilation of diverse endophenotypes, or intermediate phenotypes, such as fibrosis and thrombosis, that have a genetic association and serve as disease biomarkers or risk factors (5).

Heterogeneity in cardiovascular disease phenotypes is recognized further via the involvement of numerous cell types and complex combinations of interacting molecular species, manifestation of interrelated disease features, and the expression of variable clinical trajectories. This heterogeneity requires important considerations that are unique to cardiovascular medicine when information is extrapolated from big data. Importantly, the pathobiology of cardiovascular diseases often stands in stark contrast to diseases whose expression is tightly coupled to a sentinel molecular event, such as phenotype switching of lymphoid cells in leukemia (6), or coupled to an acquired but reversible insult driving a pathophenotype, such as extreme nutritional deficiency syndromes (7).

In this Review, we propose key strategies to improve the application of big data in cardiovascular disease by clarifying the relationships among genotype, endophenotype, and clinical phenotype. We discuss examples that show the true translational potential of unbiased data sets, and also offer cutting-edge analytical approaches to address wide-ranging limitations in the conventional reductionist path toward precision medicine.

Deconstructing genotype as the determinant of cardiovascular phenotype

The common viewpoint that most cardiovascular diseases result from a heritable component (or a limited number of heritable components) has fueled the search for pathogenic genes to implicate as causal factors. Certainly, there are notable examples of single gene mutations that are directly causal for cardiovascular disease, such as the relationship among mutations in the low-density lipoprotein receptor (LDLR) gene, familial hypercholesterolemia, and (accelerated) atherosclerosis (8). There are also genetic variants that are protective against coronary heart disease, including...
protein-inactivating variants in the sterol transporter NPC1L1, which correlate with improved outcome in patients treated with anti-NPC1L1 pharmacotherapy (9, 10); and loss-of-function mutations in ANGPTL4, which are associated with lower triglyceride levels and protection from cardiovascular disease (11, 12). This reductionist approach to defining a complex genotype-phenotype relationship, however, is implausible for most cardiovascular diseases with diverse and nuanced phenotypic features (13). Reductionism in medicine purports that a pathogenetic variant functions as the principal determinant of a disease trait or endophenotype, which, in turn, is a critical step in the development of a clinical disorder. Rather, the expression of overt cardiovascular end-pathophenotypes, such as myocardial infarction (MI), more likely represents the convergence of perturbations in numerous, potentially phenotypically related genes that are subject to further modification by an individual’s exposome, the cumulative environmental exposures that affect health over an individual’s lifespan (14).

As with many other diseases, large-scale GWAS represent a major strategy for identifying and implicating pathogenic genes in cardiovascular diseases. This approach, however, has inherent limitations for discovery, as GWAS are only able to provide an association between gene regions harboring the pathogenic gene and the disease phenotype, and marginal population sizes (even for highly prevalent cardiovascular diseases) limit the statistical power needed to seek even simple gene-gene interactions. Furthermore, although coverage of the genome continues to improve, restrictions on sensitivity are imposed by the depth and coverage of the sequencing platform. GWAS used to dissect complex disorders are also limited by the commonality of the genetic heterogeneity of many variants, as recently demonstrated by a GWAS meta-analysis of 184,305 individuals with MI and referents (15). The Exome Aggregation Consortium further established that, on average, individuals carried 54 mutations considered pathogenic, but approximately 41 of these mutations occurred with high enough frequency in the general population that they were considered unlikely to be causal for severe disease (16).

In contradistinction to pursuing identification of disease genes related to cardiovascular phenotypes, an alternative approach has focused on identifying genes associated with stress adaptation or resilience that are protective against adverse cardiovascular phenotypes. The existence of resilience genes or gene modifiers is substantiated by the identification of individuals with highly penetrant disease-causing mutations that do not manifest the disease phenotype, as occurs in heritable pulmonary arterial hypertension (PAH) and hypertrophic cardiomyopathy (HCM) (17–19). An individual’s exposome also modulates resilience. Investigators developed a polygenic risk score derived from 50 SNPs with genome-wide significance to test its utility in predicting coronary artery disease (CAD). When assessed in a cohort of 55,685 participants, individuals with a high genetic risk score who subscribed to three of four healthy lifestyle factors (no current tobacco use, no obesity, regular exercise, healthy diet) had a 46% reduction in relative risk for coronary events (20). This study and other similar studies using GWAS data to explore polygenic diseases represent a straightforward effort to define genomic context as the basis for disease heterogeneity. Yet, despite their improvement over simple GWAS, these studies purely remain associative and are severely limited by imperfect and incomplete data sets for the populations studied. Here, big data has the opportunity to provide enhanced clarity between genomics and precision phenotyping by increasing sample size, enhancing the characterization of more nuanced pathophenotypes, providing data important for understanding gene modification(s), and offering unique analytical strategies heretofore unavailable, chief among which is the application of network medicine to complex, heterogeneous molecular interactions (see below) (Figure 1).

Deconstructing clinical cardiovascular phenotypes

Most cardiovascular diseases are not circumspect entities that involve a single cell type or organ system in isolation. This principle has important implications for the applicability of a conventional reductionist model in informing an understanding of cardiovascular disease pathogenesis and phenotype. In addition, focusing on single genetic, molecular, or biological features based on expression frequency (or magnitude) alone does not necessarily elucidate functionality and ignores the potential for exploring pleiotropic heterogeneity in disease mechanisms and expression (21). Moreover, select cardiovascular phenotypes are identifiable only following a challenge, such as the use of exercise to provoke myocardial ischemia or in clinical trials in which the responsiveness to intervention defines the phenotype. This notion has wider implications for determining pathogenicity using big data in general and for contextualizing the genotype-endophenotype–clinical phenotype (GECP) relationship in cardiovascular diseases in particular. Here, we present selected examples of complex cardiovascular pathophenotypes that are unlikely to be reduced to a simplistic GECP relationship, and illustrate the dilemma that big data analytics must address.

**Myocardial infarction and ischemic heart disease.** The acute, late, and chronic phases of MI are associated with stressors that trigger crosstalk between cardiomyocytes and fibroblasts (22, 23), endothelial cells (24), and circulating immune cells, such as monocytes and T lymphocytes (25). The result is a mixed ischemic cardiomyopathy clinical phenotype with variable extent of dysregulated cardiomyocyte energetics, myocardial thinning, and replacement fibrosis. These changes, in turn, correspond to differences in systolic function, myocardial remodeling, and mechanical complications across different patients and temporally within the same patient (26). In considering the totality of these events, prediction of outcome based on cardiac morphology and clinical events in ischemic heart disease remains challenging and, from first principles, would seem to involve more than a single or few predisposing genetic risk factors.

**Hypertrophic cardiomyopathy.** HCM has long been viewed as a monogenic disease owing to the discovery of over 2000 variants in at least 11 genes encoding proteins of the cardiac sarcomere in affected patients (27). The rate of nonsynonymous sarcomere variants in “HCM genes” in population studies, however, predicts a disease prevalence that is approximately 2.5-fold greater than is observed clinically in large echocardiographic studies (28, 29). Analyses leveraging exome and whole-genome databases that are inclusive of large normal populations, resources more readily available only in recent years, demonstrate that some HCM mutations considered causative are observed at different frequencies in com-
structure and function attributable to endothelial dysfunction (34), increased vascular smooth muscle stiffness (35), aberrant matrix production by adventitial fibroblasts and resident progenitor cells (36), and inflammatory/immune cell infiltration of vessels (37). Phenotypic heterogeneity is compounded further by the fact that hypertension is provoked by physiological stress, neurohormonal regulators, salt intake, and obesity, which may occur alone or in concert (38). While efforts to identify a mono- or polygenic basis for hypertension have identified variants associated with the pathophenotypes, initial positive observations typically have not been replicated within populations of a similar ethnic background. The Framingham Heart Study identified 33 SNPs related to blood pressure or hypertension, demonstrated that the prevalence of hypertension increased commensurate with genetic risk, and validated the SNPs by confirming genome-wide significance in 34,433 individuals (39). Similarly, GWAS of 140,886 individuals of European ancestry enrolled in the UK Biobank identified 107 loci related to blood pressure traits. More recently, a GWAS meta-analysis of over 1 million individuals identified 535 new loci linked to blood pressure traits (40). Despite these efforts, the top candi-
adventitial fibroblasts contribute to PAH endophenotypes, and that increased oxidant stress, dysregulated metabolism, apoptosis resistance, cell proliferation, and fibrosis that underlie vascular lesions in PAH have been reported in patients without aberrant BMPR2 signaling (44, 45).

Concept of converging pathophenotypes. Many cardiovascular diseases exhibit broad phenotypic heterogeneity, with manifestation of clinical disease comprising a spectrum of potentially related subphenotypes that converge on a common end-pathophenotype (Figure 2). Deconstruction of these converging cardiovascular phenotypes at the disease expression level is an area that will likely be informed meaningfully by big data. Disease heterogeneity is evident in CAD, in which the clinical manifestation (stable CAD versus acute coronary syndromes versus acute MI), underlying anatomic pathology (lipid-rich thin cap fibroatheroma versus fibrotic negative remodeling), and molecular mechanisms that contribute to disease pathogenesis are all heterogeneous processes that may be operative alone or simultaneously (46). HF with preserved ejection fraction is also recognized to have substantial underlying heterogeneity, which was revealed through clinical phenomapping and unbiased clustering analysis. A cohort of patients who appeared phenotypically similar were clustered into three groups on the basis of clinical characteristics with appreciable differences in the risk of HF hospitalization (HR, 4.2; 95% CI, 2.0–9.1; P < 0.001) (47). Phenotypic heterogeneity has also been described for PAH with evidence of differential responses to acute challenge with vasodilators, and for cardiomyopathies (48, 49).

Thus, common cardiovascular diseases are not phenotypically homogeneous entities, but, rather, an assembly of widely diverse endophenotypes leading to the panoply of clinical phenotypes among individual patients and within populations.

**Phenotypic heterogeneity in cardiovascular diseases and clinical trials.** Big data is well positioned to resolve heterogeneous clinical phenotypes by bridging the divide between a reductionist approach to cardiovascular phenotypes and the reconstruction of a high-fidelity, multifaceted GECP relationship. Data mining from the Framingham Heart Study and the Nurses’ Health Study,
strating response to drug. Drug nonresponders were found to har- 
bor functional variants in \(SIRT3\) and \(UCP2\) (63). As yet another 
example, endophenotypic and phenotypic heterogeneity is pre -
sumed to be responsible, in part, for spironolactone’s failure to 
reduce the incidence of a composite endpoint of cardiovascular 
death and adverse cardiac events in patients with HF and pre-
served ejection fraction (64), which contrasts with the benefits 
obscerved for this drug in patients with HF and reduced ejection 
fraction (64, 65) (although the possibility that trial compliance 
among patients in ref. 64 may have varied by geographic region 
needs also to be considered).

Phenotypic heterogeneity in clinical trials is also imparted 
by temporal changes in definition of what constitutes cardio-
vascular disease. Thresholds for categorizing a phenotype as 
healthy, at-risk, or disease are subject to reclassification based 
on accumulating evidence. This reclassification has occurred in 
recent years for hypertension (66), hypercholesterolemia (67), 
and pulmonary hypertension (68), and has substantial implica-
tions for how big data manages historical data recorded using 
previously accepted clinical benchmarks.

Modernizing the approach to understanding endophenotypes. 
Advancing technological capabilities continue to improve the 
throughput and depth of genomic, metabolomic, proteomic, and 
transcriptomic profiling assays. Next-generation tools on the hori-
zon, such as nanopore technology, in situ nucleic acid sequenc-
ing, and enhanced molecular imaging platforms, promise a new 
dimension of high-resolution data collection with greater trans-
lational potential compared with current standards (69). These
exciting advances, nonetheless, seem likely to build upon a pervasive problem in the interpretation of big data relative to precision medicine: providing context to results. For example, findings from state-of-the-art RNA-Seq methods now allow mapping of reads with exquisite accuracy and reproducibility, although personalization of the meaning of outputs is unlikely to be resolved by a simple rank ordering of differentially expressed genes alone, which is currently a standard approach (70). The complexity of this challenge escalates when multiple omics-based platforms are integrated and these data are coupled with clinical descriptors mined from the EHR. Here, then, we propose three strategies for addressing this pervasive and growing problem.

First, acknowledging that limitations of the current GECP model welcome new paradigms, we propose that most cardiovascular diseases are complex and involve multiple overlapping endophenotypes (e.g., fibrosis, thrombosis, inflammation, apoptosis resistance, calcification) that converge to determine a specific clinical disorder (71). This approach reorganizes the genotype-endophenotype relationship from a model of reductive divergence, in which a specific mutation is responsible for all disease features, to one of convergence (Figure 3). This alternative approach provides greater flexibility for integrating genetic risk, environmental triggers or modifiers, and crosstalk between molecular pathways, such as protein-protein interactions (PPIs), as the basis of cardiovascular (and all other) diseases. In this model, individual endophenotypes are regulated by a network of PPIs, and critical proteins in this disease network that are modified by genotype, acquired factors, exposures, or a combination thereof serve to individualize the clinically observed pathophenotype (e.g., fibrosis-dominant HCM, calcification-dominant CAD) and, in doing so, may also provide novel insights into disease inception.

Second, perinatal, developmental, and epigenetic determinants of biological makeup are often overlooked despite reproducible data supporting their importance for diseases of adulthood. In adults, a history of very low preterm birthweight is associated with a 40% increase in 2-hour insulin concentration following a standard glucose challenge and a 4.8-mmHg increase in systolic blood pressure and correlates positively with CAD incidence (72, 73). These observations have been expanded more recently to include a positive association between preterm birth status and adult-onset pulmonary hypertension, among other cardiovascular diseases (74, 75). Alterations to the normative, predictive fetal adaptation response during the transition from the prenatal to the postnatal environment, particularly metabolic reprogramming as well as more specific epigenetic mechanisms, have been proposed to account for these observations (76). It may be the case that unraveling variance in metabolomic data in adult cardiovascular diseases requires greater consideration of the developmental origins of disease, including transgenerational epigenetic factors from grandparents (77, 78). Although perinatal information may not be readily available at point of care, it should be noted that family history itself is generally underutilized in the process of individualizing the GECP relationship of a given patient.

Third, incorporation of data from validated personal health monitoring devices, including daily exercise dose, physiological (e.g., sleep) parameters, and detailed nutritional data, is likely to prove pivotal for clarifying acquired cues that inform individualized cardiovascular profiles. Integrating data collected from various biospecimens is also an emerging strategy for refining the interpretation of endophenotypes (79), as insights from the thiol redox metabolome in saliva and urine (80), gas chromatographic analysis of expired volatile biomarkers (81), and the human-microbiome relationship (82) have already advanced knowledge on atherosclerotic vascular disease, MI, and stroke (83, 84).

Limitations and opportunities in using big data for the future of cardiovascular medicine

Limitations: identifying current roadblocks as avenues for future innovation. The advent of next-generation high-throughput omics technologies has allowed for deep molecular phenotyping of disease tissue concomitant with precision clinical phenotyping. Accessibility of disease tissue, and now liquid biopsy (83), formed the foundation for precision phenotyping in cancer by vertical integration of data from genomics to clinical laboratory and imaging results to outcomes. This pipeline is viable, however, in only a limited number of cardiovascular diseases owing to the absence of routinely available cardiac and vascular tissue. In current practice, endomyocardial biopsy to access right ventricular tissue is limited mainly to transplant cardiology or new-onset fulminant HF where myocarditis is suspected, while left ventricular biopsy is rarely performed (86). Liquid biopsy of the heart, however, is already in routine use and is performed to detect myocardial injury and MI through the well-known measurement of troponin levels. The diagnostic and prognostic utility of liquid biopsy as a long-term solution for precision endophenotyping hinges on the ability to detect organ- and disease-relevant circulating cells, circulating organ-specific biomarkers such as microRNAs, and cell-free DNA that have a genomic profile that differentiates between cardiovascular health and disease (87, 88). Without access to disease tissue or relevant surrogates, accurate and informative detailed endophenotyping will remain incomplete.

The EHR is recognized as a repository of diverse longitudinal data sets that inform cardiovascular disease and incorporate laboratory, imaging, biological, and descriptive data with heterogeneity in data collection frequency (89). As such, the EHR is primed for data mining. Despite this wealth of information, the EHR lacks standardization and a global universal language to facilitate interoperability between investigators and across different platforms. The ability to harvest all data variables, including those not believed to be associated with the (patho)phenotype of interest (“orthogonal” features), from the EHR also depends on the depth of the analytic tools available, and is subject to the same data quality, reliability, and inconsistency issues observed with any large-scale data source (reviewed in ref. 90). Methodologies have been introduced to validate error-prone EHRs, and it is likely that these will evolve as the data sources become increasingly large and complex (91). Thus, assimilating EHR data into precision phenotyping and cardiovascular care remains a key, ongoing challenge.

The increasing size of EHR data sets further necessitates the use of high-performance computing (supercomputing with parallel processing) with a move toward exascale computing: a computational system that can perform a quintillion 10 18 calculations per second (92, 93). Such high-performance computing will accelerate and facilitate the use of machine learning and artificial (auxiliary)
intelligence (AI) in clinical medicine in general (94, 95), and cardiovascular medicine in particular (96). Among its many useful applications from a big data perspective, AI has informed cardiovascular medical imaging analyses and enhanced phenotyping; clinical decision making and risk prediction; identification of novel phenotype clusters or cohorts; and genomic-phenomic analyses of complex data sets in which previously undisclosed relationships (some causative) may be revealed.

Although landmark clinical trials have informed endophenotyping and phenotyping in cardiovascular disease, continued demonstration of disease heterogeneity requires that the GECP relationship be continually refined. Available resources for improving precision phenotyping, however, are limited by what would now be considered incomplete or imprecise data from older sources or concluded trials, the inability to gather new data from these resources, and the absence of biospecimens to perform new or updated omics testing. Thus, historical studies and data sets may have limited applicability for future analyses. While big data analytics may overcome this limitation by compiling disparate data from a large-sized sample collective, this challenge also underscores the need to consider the future relevance of data and biospecimens collected in clinical trial design and the creation of large-scale inclusive and integrated studies at a (multi)national level.

Opportunities: clinical and integrated biological-clinical networks. Clinical networks, in which nodes and links are represented by physiological parameters and physiological effects, respectively, have been reported on a small scale (97). Although comprehensive information on a larger “physiome” remains a goal, a modified approach using correlative networks has already proven useful in numerous venues, including comorbidity-driven schemata that improve pediatric cardiovascular disease diagnostics, and reports that clarify the clinical heterogeneity of patients with chronic obstructive pulmonary disease, a common cardiovascular disease comorbidity (98, 99).

This approach may be particularly helpful in deciphering cardiovascular diagnostic testing results, which often rely on branch chain logic to organize data. Such methods may overlook collections of interrelated variables that inform nuanced pathophysiological or are effective for identifying patient subgroups. For example, invasive cardiopulmonary exercise testing is a comprehensive test used to diagnose unexplained dyspnea and generates approximately 100 measurements per patient that span seven physiological parameters (e.g., central cardiopulmonary hemodynamics, respiratory gas exchange, and others). Current methods used to interpret these tests, however, typically focus on a very small subset of information, usually fewer than five variables (100). An exercise network was reported recently that included 39 nodes and 98 links, providing comprehensive information on unexpected relationships between testing measurements that included variables across numerous different exercise parameters (e.g., lung function, right ventricular function). This network was reduced further to a group of ten variables, which, in turn, was effective for identifying four distinct patient subgroups defined by unique clinical, exercise, and outcome profiles. From this approach, a risk prediction model was assembled that was based on network medicine, and emerged as superior to probabilistic (traditional) linear regression methods for risk stratification (101).

These and other novel approaches to classifying patients that are based principally on a collection of related clinical parameters, as reported previously for HF with preserved ejection fraction (47), may be important for phenotype selection in forward-thinking approaches to cardiovascular medicine clinical trial design. These features include enrollment criteria based on Mendelian randomization, quantitative trait loci, and adaptive trial designs using systems pharmacology–based methods that permit flexibility in patient enrollment, data collection schedule, and endpoint selection matched to the ongoing collection of data across clinical and pharmacological fronts (102). Some of these approaches have already been proposed or considered in studies on lipid-lowering agents (103), and in rare cardiovascular diseases for which the “N-of-1 trial” may ultimately prove useful for patients with a specific biological profile matching a specific pharmacotherapeutic agent (104).

Big data’s progress in cardiovascular medicine: network medicine

Network medicine can be a useful analytical approach that combines unique genomic features with unique clinical (endo)phenotypes in a fully integrated way. For example, if one considers the universe of (physical) PPI (the interactome) as a global network template, we have shown that each disease has a unique discrete subnetwork (module) within it (105). Each patient with this disease, in turn, can be analyzed for genetic variants or differentially expressed genes (proteins) in this disease module, rendering the individualized disease module or reticulotype (106). Exploring this complex personalized module for functional variation provides a pathway for personalized precision medicine designed to restore (normal) network function, correct the reticulotype, and improve the clinical (endo)phenotype (ref. 106 and Figure 3).

Network medicine can also be a useful analytical approach for clarifying the molecular mechanisms that distinguish functional subtypes within a specific endophenotype (71), each of which has its own unique module in the PPI. For example, we recently developed an endophenotype network that regulates fibrosis (the fibrosome) in which we included PPIs stratified by differing collagen biofunctionalities. We used wound healing and PAH as clinical correlates representing adaptive versus pathogenic fibrosis, respectively (45). The network was refined to focus on PPIs regulated by the pro-oxidant and profibrotic hormone aldosterone, which is implicated in both fibrosis subtypes (107, 108). Betweenness centrality, a network measure of node importance, identified the Cas protein NEDD9 as important in the phenotype transition between adaptive and pathogenic fibrosis in silico. Oxidative post-translational modification of NEDD9 at Cys46 emerged as a novel molecular mechanism that regulates pathogenic collagen synthesis with implications for PAH clinically. Overall, this line of investigation illustrates the importance of using unbiased but informed analytical methods (e.g., network medicine) to characterize the GECP relationship in a more nuanced, holistic way, devoid of the limitations of conventional reductionism (45).

Diversity in post-transcriptional mechanisms across endophenotypes has also been reported in PAH, beginning with our analysis of microRNA networks in PAH (109), wherein miR-21 was shown to regulate pathogenic signaling in the disease. Subsequently,
the miR-130/301 family has been shown to affect PPI pathways involved in inflammation, vasomotor tone, apoptosis, and hypoxia responses (110). Others have shown that miR-34a-3p regulates mitotic fission, drawing an important connection between epi-
genetics and dysregulated metabolism. Indeed, the tendency for these and other microRNAs to induce a particular vas-
cular morphological feature may be influenced by endothelin-1, other vasoactive hormones, or hypoxia (111). Additional empir-
cal data are needed, however, to establish the framework through which microRNAs or other post-transcriptional events interact with genetic risk factors to regulate complex disease features, for instance, plexovascular lesions in PAH.

Using biological data to discern specific patient subgroups from an otherwise heterogeneous clinical population is an evolv-
ing next step toward personalized medicine. Unsupervised anal-
yses of plasma proteomic data from PAH patients reinforced the possibility of a common biological thread across patients as deter-
mined by protein clusters (from a k-means analysis) (112). These clusters were enriched for different inflammatory/immune path-
ways that were not affected by patient pulmonary vascular dis-
ease subtype or comorbidities, but corresponded to differences in clinical risk. Cluster assignment itself was not determined a priori by functional relationships between proteins from the perspective of the (functional) interactome, and, thus, additional opportunity may exist to refine further this approach for optimizing biological classification of patients. Such complementary and alternative methodologies are proposed in the NIH-sponsored Pulmonary Vascular Disease Phenomics Study (PVDOMICS), which aims to integrate multidimensional omics, clinical, and outcome data from a large cohort of pulmonary hypertension patients using informed and agnostic approaches (113).

Conclusions
Cardiovascular diseases are complex heterogeneous pathopheno-
types that cannot typically be resolved by the reductionist concept of a singular GECP relationship. At each level, diversity in genotype, reticulotype, and endophenotype expression owing to modifying factors, such as the exposome, dictates clinical phenotypes, which themselves are heterogeneous. This concept is supported by the fact that it is the exception, not the rule, that genetic variants identified as disease-causing translate into a universal blueprint for an endophenotype and a specific cardiovascular disease. Increasingly large compendia of clinical trial data and matched omics data have not yet provided clarity for precision endophenotyping and clinical cardio-
vascular pathophenotypes, suggesting that we have grossly underes-
imated the sample size, data types, and analytics required to unravel heterogeneity in cardiovascular disease. Big data coupled with novel analytical approaches, such as network analyses, will have the capaci-
ty to elucidate origins of heterogeneity in cardiovascular diseases and provide clarity to the genotype–endophenotype–cardiovascular disease relationship as espoused by network medicine (106, 114).

Acknowledgments
This work was supported, in part, by NIH grant HL125215 and American Heart Association grant 19A1ML34980000 to JL; NIH grants HL131787, HL139613, and HL145420 and the National Scleroderma Foundation to BAM; and NIH grants HL061795, HL119145, HG007690, and GM107618 and American Heart Association grant D700382 to JL.

Address correspondence to: Joseph Loscalzo, Brigham and Wom-
en’s Hospital, 75 Francis Street, Boston, Massachusetts 02115, USA.
Phone: 617.732.6340; Email: jloscalzo@rics.bwh.harvard.edu.

7. Cheng Y, et al. LNK/SH2B3 regulates IL-7 receptor signaling in normal and malignant B-progeni-
8. Lehman MA, Goldstein JL, Brown MS, Russell DW, Schneider WJ. Internalization-defective LDL receptors produced by genes with nonsense and frameshift mutations that truncate the cyto-
13. Leopold JA, Loscalzo J. Emerging role of preci-
14. Kathiresan S, Srivastava D. Genetics of human car-
18. West J, et al. Gene expression in BMP2 mutation carriers and without evidence of pulmonary arterial hypertension suggests pathways relev-
19. Vermeer AMC, Clur SB, Blom NA, Wilde AAM, Christiansen I. Penetration of hypertrophic cardio-
21. Webb TR, et al. Systematic evaluation of plei-
22. Rose BA, et al. Cardiac myocyte p38 kinase reg-
23. Nagaraju CK, et al. Myofibroblast phenotype and reversibility of fibrosis in patients with end-


the field of pulmonary hypertension. Focusing on
developmental origins and disease inception for
the prevention of pulmonary hypertension. *Am J
76. Tobi EW, et al. DNA methylation of IGF2, GNAS,
INSIGF and LEP and being born small for
77. Lewis GD, et al. Metabolic profiling of right ven-
tricular-pulmonary vascular function reveals cir-
culating biomarkers of pulmonary hypertension.
78. Kaati G, Bygren LO, Edvinsson S. Cardiovascular
and diabetes mortality determined by nutrition
during parents’ and grandparents’ slow growth
79. Sharma A, et al. Integration of molecular interac-
tome and targeted interaction analysis to iden-
2018;8(1):14439.
spectrometry platform for comprehensive assess-
ment of the thiol redox metabolome. *Redox Biol*.
81. Otterbein LE, et al. Carbon monoxide has
anti-inflammatory effects involving the mito-
gen-activated protein kinase pathway. *Nat Med*.
82. Wang Z, et al. Gut flora metabolism of phospha-
tidylylcholine promotes cardiovascular disease.
is associated with an altered gut metagenome.
84. Feng Q, et al. Integrated metabolomics and
metagenomics analysis of plasma and urine
identified microbial metabolites associated with
85. Rossi G, Ignatiadis M. Promises and pitfalls of
using liquid biopsy for precision medicine.
86. Francis R, Lewis C. Myocardial biopsy: techniques
87. Jakob P, et al. Profiling and validation of
circulating microRNAs for cardiovascular
events in patients presenting with ST-segment
elevation myocardial infarction. *Eur Heart J*.
88. Xie J, Yang J, Hu P. Correlations of circulating
cell-free DNA with clinical manifestations
89. Wu PY, Cheng CW, Kaddi CD, Venugopalan
J, Hoffman R, Wang MD. –Omic and elec-
tronic health record big data analytics for
90. Rumsfeld JS, Joyst KE, Maddox TM. Big data
analytics to improve cardiovascular care:
promise and challenges. *Nat Rev Cardiol*.
91. Ouyang L, Apley DW, Mehrutra S. A design of
experiments approach to validation sampling
for logistic regression modeling with error-
prone medical records. *J Am Med Inform Assoc*.
2016;23(1):e71–e78.
92. Hoekstra AG, Chopard B, Coster D, Portegies
Zwart S, Coveney PV. Multiscale computing for
science and engineering in the era of exascale
2019;377(2124):20180144.
93. Cirillo D, Valencia A. Big data analytics for
personalized medicine. *Curr Opin Biotechnol*.
94. Vogel AL, Kvedar JC. Artificial intelligence pow-
95. Topol EJ. High-performance medicine: the con-
vergence of human and artificial intelligence.
96. Sardar P, Abbott JD, Kundu A, Aronow HD,
Granada JF, Giri J. Impact of artificial intel-
ligence on interventional cardiology: from
decision-making aid to advanced interventional
procedure assistance. *JACC Cardiovasc Interv*.
97. Guyton AC, Coleman TG, Granger HJ. Circu-
98. Ong MS, et al. Learning a comorbidity-driven
taxonomy of pediatric pulmonary hypertension.
among subjects with mild airflow obstruction in
100. Maron BA, Cockrill BA, Waxman AB, Systrom
DM. The invasive cardiopulmonary exercise test.
101. Oldham WM, et al. Network analysis to risk strat-
ify patients with exercise intolerance. *Circ Res*.
102. Antman E, Weiss S, Loscalzo J. Systems phar-
macology, pharmacogenetics, and clinical trial
design in network medicine. *Wiley Interdiscip Rev
103. Holmes MV, et al. Mendelian randomization of
blood lipids for coronary heart disease. *Europ J
104. Spiekerkoetter E, et al. Randomised placebo-
controlled safety and tolerability trial of FK506
(tacrolimus) for pulmonary arterial hyperten-
ing disease-disease relationships through the
106. Lee LV, Loscalzo J. Network medicine in pathobi-
107. Maron BA, Opotowsky AR, Landzberg MJ,
Loscalzo J, Waxman AB, Leopold JA. Plasma
aldosterone levels are elevated in patients with
pulmonary arterial hypertension in the absence
of left ventricular heart failure: a pilot study. *Europ J
reactivity by decreasing glucose-6-phosphate dehy-
pathogenic signaling to control pulmonary
hypertension: results of a network bioinformatics
110. Bertero T, et al. The microRNA-130/301 family
controls vasoconstriction in pulmonary hyper-
111. Chen KH, et al. Epigenetic dysregulation of the
dynamin-related protein 1 binding partners
MiD49 and MiD51 increases mitotic mitochon-
drial fission and promotes pulmonary arterial
hypertension: mechanistic and therapeutic
112. Sweat AJ, et al. Discovery of distinct immune
phenotypes using machine learning in pul-
monary arterial hypertension. *Circ Res*.
study to improve understanding of pulmonary
vascular disease through phenomics. *Circ Res*.
2017;121(10):1136–1139.
114. Loscalzo J, Barabasi AL, Silverman EK. *Network
Medicine: Complex Systems in Human Disease and
Therapeutics*. Cambridge, Massachusetts, USA:
Harvard University Press; 2017.