

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.

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.

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 (endo)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

Conflict of interest: JAL, BAM, and JL are inventors on a pending patent (US patent 9,605,047). JAL is an inventor on patent applications (US provisional patent applications 62/434,565 and 61/99,754). BAM is an inventor on patent applications (US provisional patent applications 24622 and 61/99,754). JL owns equity in Ionis Pharmaceuticals, Scipher Medicine, and Leap Therapeutics, and has received honoraria from Momenta Pharmaceuticals, Broadview Ventures, Sanofi, Ionis Pharmaceuticals, Leap Therapeutics, and Applied Biomath for work as a consultant and scientific advisor.

Copyright: © 2020, American Society for Clinical Investigation.

Reference information: *J Clin Invest.* 2020;130(1):29–38.

<https://doi.org/10.1172/JCI129203>.

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-phenotypes, 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 (GECPC) relationship in cardiovascular diseases in particular. Here, we present selected examples of complex cardiovascular pathophenotypes that are unlikely to be reduced to a simplistic GECPC 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 diseases 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-

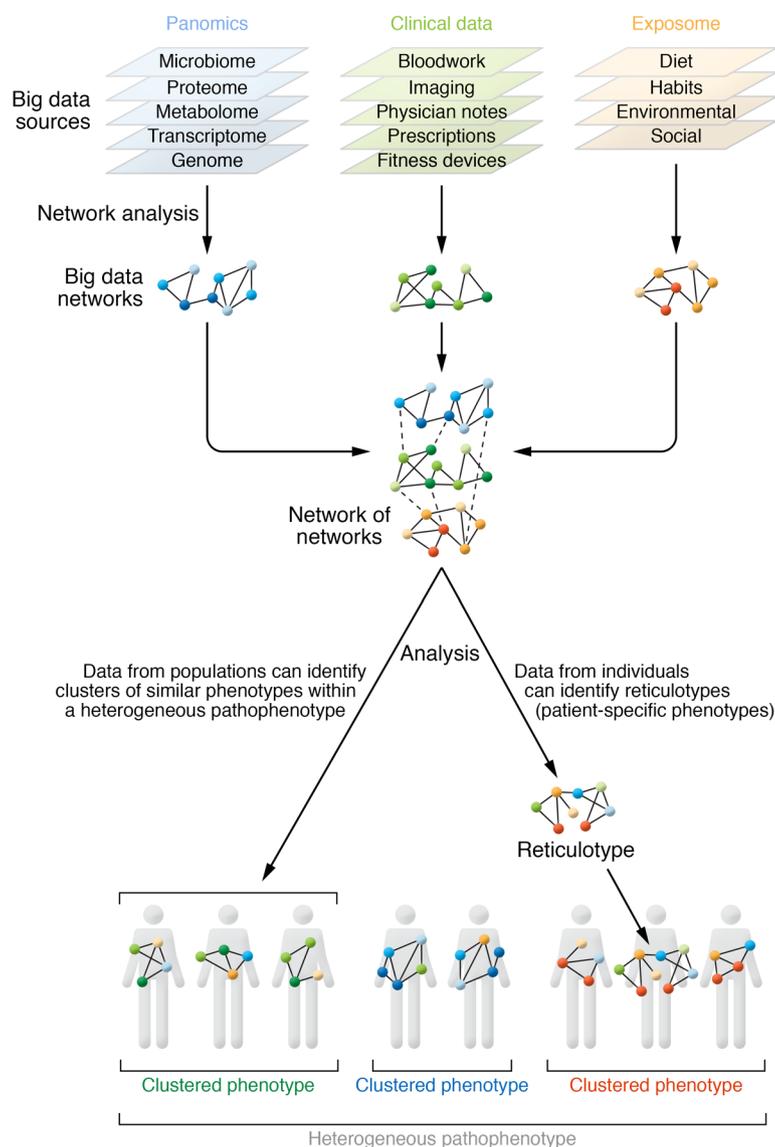


Figure 1. Big data enhances precision cardiovascular phenotyping. Contemporary understanding of the heterogeneity in cardiovascular disease requires compilation of a diverse array of big data sources. Data from these domains are amenable to novel network medicine analytics to generate individual patient networks to define networks based on population-level data as well as a reticulotype (i.e., a patient's unique molecular network that allows an exploration of how perturbations affect phenotype). While precision phenotyping may define clusters of patients, reticulotyping provides further resolution to clusters by identifying the molecular (network) drivers of unique patient-specific characteristics.

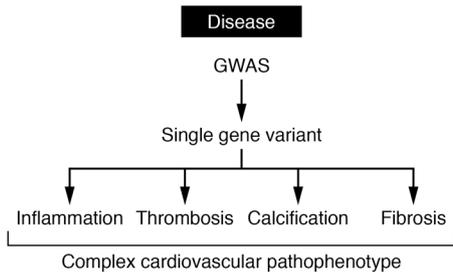
parison with nondisease controls in certain ethnic (racial) groups, and do not associate with the pathophenotype in these subgroups (30). Furthermore, the HCM spectrum spans a diverse collection of endophenotypes that do not hinge on sarcomere-dependent pathobiology. For example, mitral valve elongation, myocardial replacement fibrosis, and hypertrophic remodeling of intramural coronary arteries, among other abnormalities that involve cell types that do not express cardiomyocyte sarcomere proteins, are all observed to varying degrees in individual HCM patients (31).

Other cardiomyopathies. Missense variants in *TTN*, encoding titin, have been implicated recently in the pathogenesis of left ventricular (LV) dilated cardiomyopathy and LV noncompaction cardiomyopathy (32, 33). The range of titinopathies is, therefore, vast, but the extent to which each variant is disease-causing is uncertain, particularly in LV noncompaction. This entity is characterized by particularly heterogeneous structural features, and often tracks with other congenital anatomic abnormalities that are not likely to depend directly on titin.

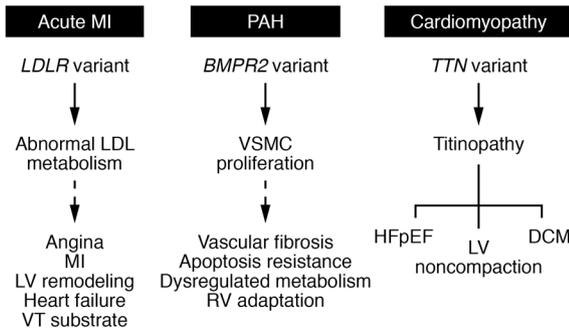
Hypertension. There is considerable heterogeneity in the hypertensive pathophenotype, with dysregulation of vascular

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-

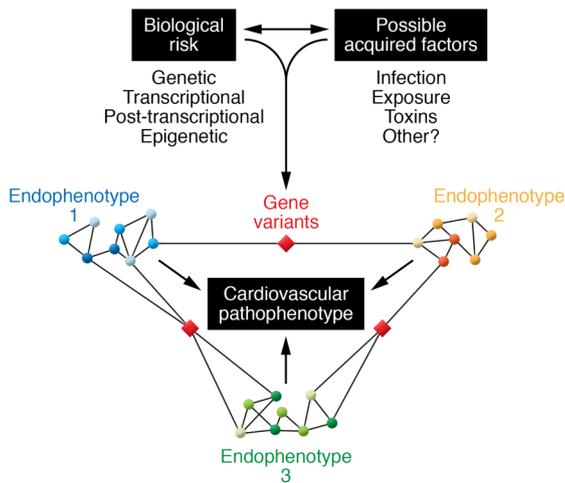
A Classic model



B Examples of cardiovascular disease heterogeneity



C Contemporary model



date disease genes were dissimilar among studies, and the role of gene-modifying environmental factors and other demographic features that may be differentially represented in the study populations (ethnicity, race) was not assessed (41).

Pulmonary arterial hypertension. A germline mutation in *BMPR2*, encoding bone morphogenetic protein receptor-2, accounts for 75% of hereditary PAH cases. Penetrance among carriers, however, is highly variable, ranging from 20% to 80% depending on the study population and design (42). This finding suggests that environmental cues may be required to induce the clinical phenotype in many carriers. It is also noteworthy that *BMPR2*, a member of the TGF- β superfamily of receptors, seems to modulate vascular remodeling predominantly via effects on pulmonary artery smooth muscle cell growth and survival (43). It is important to note, however, that endothelial cells, pericytes, and

Figure 2. Heterogeneity in cardiovascular disease and convergence on a common end-pathophenotype. (A) Cardiovascular diseases are complex clinical phenotypes that involve many different endophenotypes (e.g., inflammation, thrombosis, calcification, fibrosis) that cannot be explained solely by a single pathogenic variant. (B) Heterogeneity in cardiovascular diseases is evident as shown by the relationships among genetic variants (genotypes), the biochemical and cellular consequences of harboring these variants (endophenotypes), and clinically observed pathophenotypes. (C) In a model based on big data and network analyses, specific endophenotypes are determined by modules or a (sub) network of protein-protein interactions within a larger disease network. Crosstalk between pathways that regulate different endophenotypes via a critical gene may occur. In this way, post-transcriptional and epigenetic mechanisms that are important in the pathogenesis of disease endophenotypes are emphasized and, collectively, converge to produce a complex pathophenotype. DCM, dilated cardiomyopathy; HFpEF, heart failure with preserved ejection fraction; LDL, low-density lipoprotein; LV, left ventricle; MI, myocardial infarction; RV, right ventricle; VSMC, vascular smooth muscle cell; VT, ventricular tachycardia. Adapted with permission from the *Journal of the American College of Cardiology* (network image in Figure 2C of, and bottom right panel of central illustration of, ref. 31).

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 GEC relationship. Data mining from the Framingham Heart Study and the Nurses' Health Study,

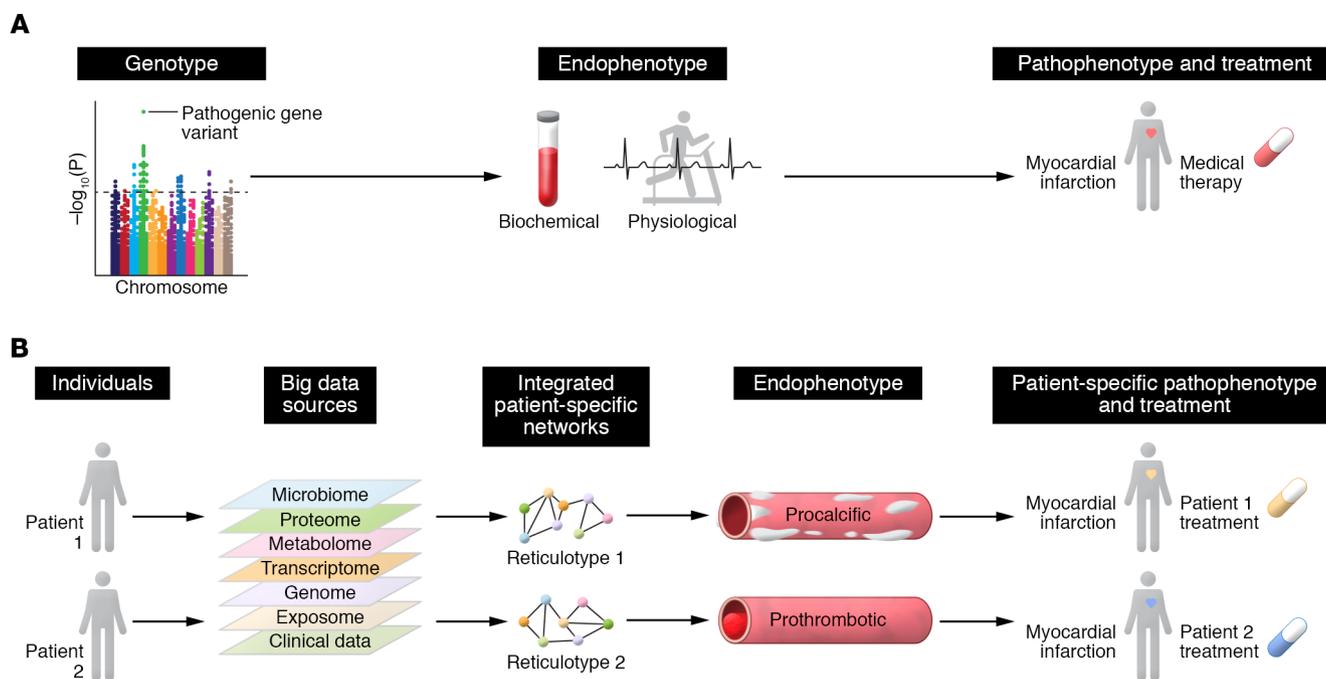


Figure 3. Big data informs reticulotyping and cardiovascular disease phenotyping. (A) The current viewpoint of cardiovascular disease phenotyping focuses on reductionism, which posits that a pathogenic variant is causal for a disease trait, or endophenotype, and, therefore, a key determinant of developing a cardiovascular disease. (B) Network medicine allows for precision endophenotyping and phenotyping for individuals with similar clinical signs and symptoms. Using big data, patient-specific integrated networks (e.g., protein-protein interaction networks) can be constructed, and the consequences of perturbations owing to an individual's unique genomic and molecular makeup, known as the reticulotype, can be explored. The reticulotype, in turn, also governs endophenotype and defines a patient-specific phenotype that may not have been evident previously.

the CHARGE Consortium, SWEDEHEART, and the UK Biobank Initiative, among others, has collectively explored GEC relationships for hypertension, coronary heart disease, and MI (39, 50–54). These findings will surely be advanced by future reports from even larger-scale cohorts, such as the Million Veterans Program (55) and the NIH All of Us study (56). Although big data is changing the strength of evidence available from cardiovascular observational studies, merging data from these and other increasingly large population-based registries, biobanks, and electronic health records (EHRs) remains a substantial challenge. As such, big data is more likely to be a facet incorporated into clinical trial design rather than a replacement for randomized clinical trials designed to establish causal effects for cardiovascular diseases (57).

Endophenotypic heterogeneity as the driving factor in phenotypic heterogeneity has been recognized and explored in few clinical trials, most notably the association between *CYP2C19* loss-of-function allele(s), diminished platelet inhibition by clopidogrel in carriers, and cardiovascular or cerebrovascular events (58–60). A rational genotype-driven treatment strategy, however, has not always translated into superior efficacy or safety, as demonstrated by studies that pursued genotyping of *CYP2C9* or of *VKORC1* to guide warfarin dosing (61, 62). Conversely, if endophenotypic or phenotypic heterogeneity is overlooked in enrolling subjects in clinical studies, such heterogeneity often emerges as a contributor to unexpected outcomes. For example, despite abundant supportive preclinical data, an open-label study of dichloroacetate in 20 patients with idiopathic PAH on background therapy revealed substantial interindividual variability, with only 7 patients demon-

strating response to drug. Drug nonresponders were found to harbor functional variants in *SIRT3* and *UCP2* (63). As yet another example, endophenotypic and phenotypic heterogeneity is presumed 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 preserved ejection fraction (64), which contrasts with the benefits observed 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 cardiovascular 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 implications 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 horizon, such as nanopore technology, in situ nucleic acid sequencing, and enhanced molecular imaging platforms, promise a new dimension of high-resolution data collection with greater translational 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 normal, 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 individual-

ized 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 (85), 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 pathophysiology 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 Cys¹⁸ 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 epigenetics and dysregulated metabolism. Indeed, the tendency for these and other microRNA families to induce a particular vascular morphological feature may be influenced by endothelin-1, other vasoactive hormones, or hypoxia (111). Additional empirical 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, plexogenic vascular lesions in PAH.

Using biological data to discern specific patient subgroups from an otherwise heterogeneous clinical population is an evolving next step toward personalized medicine. Unsupervised analyses of plasma proteomic data from PAH patients reinforced the possibility of a common biological thread across patients as determined by protein clusters (from a *k*-means analysis) (112). These clusters were enriched for different inflammatory/immune pathways that were not affected by patient pulmonary vascular disease 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 pathophenotypes that cannot typically be resolved by the reductionist concept of a singular GECF 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 cardiovascular pathophenotypes, suggesting that we have grossly underestimated 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 capacity 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 JAL; 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 Women's Hospital, 75 Francis Street, Boston, Massachusetts 02115, USA. Phone: 617.732.6340; Email: jloscalzo@rics.bwh.harvard.edu.

- Joyner MJ, Paneth N. Promises, promises, and precision medicine. *J Clin Invest*. 2019;129(3):946–948.
- Owan TE, Hodge DO, Herges RM, Jacobsen SJ, Roger VL, Redfield MM. Trends in prevalence and outcome of heart failure with preserved ejection fraction. *N Engl J Med*. 2006;355(3):251–259.
- Paneni F, Diaz Cañestro C, Libby P, Lüscher TF, Camici GG. The aging cardiovascular system: understanding it at the cellular and clinical levels. *J Am Coll Cardiol*. 2017;69(15):1952–1967.
- Pandey A, et al. Physical activity, fitness, and obesity in heart failure with preserved ejection fraction. *JACC Heart Fail*. 2018;6(12):975–982.
- van der Ende MY, Said MA, van Veldhuisen DJ, Verweij N, van der Harst P. Genome-wide studies of heart failure and endophenotypes: lessons learned and future directions. *Cardiovasc Res*. 2018;114(9):1209–1225.
- Bennett SE, Schmitt WP, Stanford FC, Baron JM. Case 22-2018: a 64-year-old man with progressive leg weakness, recurrent falls, and anemia. *N Engl J Med*. 2018;379(3):282–289.
- Cheng Y, et al. LNK/SH2B3 regulates IL-7 receptor signaling in normal and malignant B-progenitors. *J Clin Invest*. 2016;126(4):1267–1281.
- Lehrman MA, Goldstein JL, Brown MS, Russell DW, Schneider WJ. Internalization-defective LDL receptors produced by genes with nonsense and frameshift mutations that truncate the cytoplasmic domain. *Cell*. 1985;41(3):735–743.
- Stitzel NO, et al. Myocardial Infarction Genetics Consortium I, inactivating mutations in NPC1L1 and protection from coronary heart disease. *N Engl J Med*. 2014;371(22):2072–2082.
- Cannon CP, et al. Ezetimibe added to statin therapy after acute coronary syndromes. *N Engl J Med*. 2015;372(25):2387–2397.
- Myocardial Infarction Genetics CARDIoGRAM Exome Consortia Investigators, et al. Coding variation in ANGPTL4, LPL, and SVEP1 and the risk of coronary disease. *N Engl J Med*. 2016;374(12):1134–1144.
- Dewey FE, et al. Inactivating variants in ANGPTL4 and risk of coronary artery disease. *N Engl J Med*. 2016;374(12):1123–1133.
- Leopold JA, Loscalzo J. Emerging role of precision medicine in cardiovascular disease. *Circ Res*. 2018;122(9):1302–1315.
- Kathiresan S, Srivastava D. Genetics of human cardiovascular disease. *Cell*. 2012;148(6):1242–1257.
- Nikpay M, et al. A comprehensive 1,000 Genomes-based genome-wide association meta-analysis of coronary artery disease. *Nat Genet*. 2015;47(10):1121–1130.
- Lek M, et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature*. 2016;536(7616):285–291.
- Chen R, et al. Analysis of 589,306 genomes identifies individuals resilient to severe Mendelian childhood diseases. *Nat Biotechnol*. 2016;34(5):531–538.
- West J, et al. Gene expression in BMPR2 mutation carriers with and without evidence of pulmonary arterial hypertension suggests pathways relevant to disease penetrance. *BMC Med Genomics*. 2008;1:45.
- Vermeer AMC, Clur SB, Blom NA, Wilde AAM, Christiaans I. Penetrance of hypertrophic cardiomyopathy in children who are mutation positive. *J Pediatr*. 2017;188:91–95.
- Khera AV, et al. Genetic risk, adherence to a healthy lifestyle, and coronary disease. *N Engl J Med*. 2016;375(24):2349–2358.
- Webb TR, et al. Systematic evaluation of pleiotropy identifies 6 further loci associated with coronary artery disease. *J Am Coll Cardiol*. 2017;69(7):823–836.
- Rose BA, et al. Cardiac myocyte p38 α kinase regulates angiogenesis via myocyte-endothelial cell cross-talk during stress-induced remodeling in the heart. *J Biol Chem*. 2017;292(31):12787–12800.
- Nagaraju CK, et al. Myofibroblast phenotype and reversibility of fibrosis in patients with end-stage heart failure. *J Am Coll Cardiol*. 2019;73(18):2267–2282.
- Zhang Y, et al. Endothelial cells regulate

- cardiac myocyte reorganisation through β 1-integrin signalling. *Cell Physiol Biochem*. 2015;35(5):1808–1820.
25. Heidt T, et al. Differential contribution of monocytes to heart macrophages in steady-state and after myocardial infarction. *Circ Res*. 2014;115(2):284–295.
 26. Halliday BP, Cleland JGF, Goldberger JJ, Prasad SK. Personalizing risk stratification for sudden death in dilated cardiomyopathy: the past, present, and future. *Circulation*. 2017;136(2):215–231.
 27. Toepfer CN, et al. Hypertrophic cardiomyopathy mutations in *MYBPC3* dysregulate myosin. *Sci Transl Med*. 2019;11(476):eaat1199.
 28. Bick AG, et al. Burden of rare sarcomere gene variants in the Framingham and Jackson Heart Study cohorts. *Am J Hum Genet*. 2012;91(3):513–519.
 29. Maron BJ, Gardin JM, Flack JM, Gidding SS, Kurosaki TT, Bild DE. Prevalence of hypertrophic cardiomyopathy in a general population of young adults. Echocardiographic analysis of 4111 subjects in the CARDIA Study. Coronary Artery Risk Development in (Young) Adults. *Circulation*. 1995;92(4):785–789.
 30. Manrai AK, et al. Genetic misdiagnoses and the potential for health disparities. *N Engl J Med*. 2016;375(7):655–665.
 31. Maron BJ, Maron MS, Maron BA, Loscalzo J. Moving beyond the sarcomere to explain heterogeneity in hypertrophic cardiomyopathy: JACC Review Topic of the Week. *J Am Coll Cardiol*. 2019;73(15):1978–1986.
 32. Herman DS, et al. Truncations of titin causing dilated cardiomyopathy. *N Engl J Med*. 2012;366(7):619–628.
 33. Hastings R, et al. Combination of whole genome sequencing, linkage, and functional studies implicates a missense mutation in titin as a cause of autosomal dominant cardiomyopathy with features of left ventricular noncompaction. *Circ Cardiovasc Genet*. 2016;9(5):426–435.
 34. Bleakley C, Hamilton PK, Pumb R, Harbinson M, McVeigh GE. Endothelial function in hypertension: victim or culprit? *J Clin Hypertens (Greenwich)*. 2015;17(8):651–654.
 35. Sehgel NL, et al. Increased vascular smooth muscle cell stiffness: a novel mechanism for aortic stiffness in hypertension. *Am J Physiol Heart Circ Physiol*. 2013;305(9):H1281–H1287.
 36. Wu J, et al. Origin of matrix-producing cells that contribute to aortic fibrosis in hypertension. *Hypertension*. 2016;67(2):461–468.
 37. Norlander AE, Madhur MS, Harrison DG. The immunology of hypertension. *J Exp Med*. 2018;215(1):21–33.
 38. Korner PI. The phenotypic patterns of essential hypertension are the key to identifying “high blood pressure” genes. *Physiol Res*. 2010;59(6):841–857.
 39. Levy D, et al. Genome-wide association study of blood pressure and hypertension. *Nat Genet*. 2009;41(6):677–687.
 40. Evangelou E, et al. Genetic analysis of over 1 million people identifies 535 new loci associated with blood pressure traits. *Nat Genet*. 2018;50(12):1755.
 41. Mattson DL, Liang M. Hypertension: from GWAS to functional genomics-based precision medicine. *Nat Rev Nephrol*. 2017;13(4):195–196.
 42. Larkin EK, et al. Longitudinal analysis casts doubt on the presence of genetic anticipation in heritable pulmonary arterial hypertension. *Am J Respir Crit Care Med*. 2012;186(9):892–896.
 43. Upton PD, Davies RJ, Tajsic T, Morrell NW. Transforming growth factor- β (1) represses bone morphogenetic protein-mediated Smad signaling in pulmonary artery smooth muscle cells via Smad3. *Am J Respir Cell Mol Biol*. 2013;49(6):1135–1145.
 44. Zhang H, et al. Metabolic and proliferative state of vascular adventitial fibroblasts in pulmonary hypertension is regulated through a microRNA-124/PTBP1 (polypyrimidine tract binding protein 1)/pyruvate kinase muscle axis. *Circulation*. 2017;136(25):2468–2485.
 45. Samokhin AO, et al. NEDD9 targets *COL3A1* to promote endothelial fibrosis and pulmonary arterial hypertension. *Sci Transl Med*. 2018;10(445):eaap7294.
 46. Dai X, Wiernek S, Evans JP, Runge MS. Genetics of coronary artery disease and myocardial infarction. *World J Cardiol*. 2016;8(1):1–23.
 47. Shah SJ, et al. Phenomapping for novel classification of heart failure with preserved ejection fraction. *Circulation*. 2015;131(3):269–279.
 48. Hemnes AR, et al. Peripheral blood signature of vasodilator-responsive pulmonary arterial hypertension. *Circulation*. 2015;131(4):401–409.
 49. Forleo C, et al. Targeted next-generation sequencing detects novel gene-phenotype associations and expands the mutational spectrum in cardiomyopathies. *PLoS One*. 2017;12(7):e0181842.
 50. Yu E, et al. Diet, lifestyle, biomarkers, genetic factors, and risk of cardiovascular disease in the Nurses’ Health Studies. *Am J Public Health*. 2016;106(9):1616–1623.
 51. Dehghan A, et al. Genome-wide association study for incident myocardial infarction and coronary heart disease in prospective cohort studies: the CHARGE Consortium. *PLoS One*. 2016;11(3):e0144997.
 52. Szummer K, et al. Improved outcomes in patients with ST-elevation myocardial infarction during the last 20 years are related to implementation of evidence-based treatments: experiences from the SWEDHEART registry 1995–2014. *Eur Heart J*. 2017;38(41):3056–3065.
 53. Mahmood SS, Levy D, Vasan RS, Wang TJ. The Framingham Heart Study and the epidemiology of cardiovascular disease: a historical perspective. *Lancet*. 2014;383(9921):999–1008.
 54. Thompson SG, Willeit P. UK Biobank comes of age. *Lancet*. 2015;386(9993):509–510.
 55. Gaziano JM, et al. Million Veteran Program: a mega-biobank to study genetic influences on health and disease. *J Clin Epidemiol*. 2016;70:214–223.
 56. Sankar PL, Parker LS. The precision medicine initiative’s All of Us research program: an agenda for research on its ethical, legal, and social issues. *Genet Med*. 2017;19(7):743–750.
 57. Mayo CS, Matuszak MM, Schipper MJ, Jolly S, Hayman JA, Ten Haken RK. Big data in designing clinical trials: opportunities and challenges. *Front Oncol*. 2017;7:187.
 58. Mega JL, et al. Cytochrome p-450 polymorphisms and response to clopidogrel. *N Engl J Med*. 2009;360(4):354–362.
 59. Mega JL, et al. Dosing clopidogrel based on CYP2C19 genotype and the effect on platelet reactivity in patients with stable cardiovascular disease. *JAMA*. 2011;306(20):2221–2228.
 60. Wang Y, et al. Association between CYP2C19 loss-of-function allele status and efficacy of clopidogrel for risk reduction among patients with minor stroke or transient ischemic attack. *JAMA*. 2016;316(1):70–78.
 61. Pirmohamed M, et al. A randomized trial of genotype-guided dosing of warfarin. *N Engl J Med*. 2013;369(24):2294–2303.
 62. Kimmel SE, et al. A pharmacogenetic versus a clinical algorithm for warfarin dosing. *N Engl J Med*. 2013;369(24):2283–2293.
 63. Michelakis ED, et al. Inhibition of pyruvate dehydrogenase kinase improves pulmonary arterial hypertension in genetically susceptible patients. *Sci Transl Med*. 2017;9(413):eaao4583.
 64. Pitt B, et al. Spironolactone for heart failure with preserved ejection fraction. *N Engl J Med*. 2014;370(15):1383–1392.
 65. Pitt B, et al. The effect of spironolactone on morbidity and mortality in patients with severe heart failure. Randomized Aldactone Evaluation Study Investigators. *N Engl J Med*. 1999;341(10):709–717.
 66. Whelton PK, et al. 2017 ACC/AHA/AAPA/ABC/ACPM/AGS/APhA/ASH/ASPC/NMA/PCNA Guideline for the prevention, detection, evaluation, and management of high blood pressure in adults: executive summary: a report of the American College of Cardiology/American Heart Association task force on clinical practice guidelines. *Hypertension*. 2018;71(6):1269–1324.
 67. Pencina MJ, et al. Application of new cholesterol guidelines to a population-based sample. *N Engl J Med*. 2014;370(15):1422–1431.
 68. Simonneau G, et al. Haemodynamic definitions and updated clinical classification of pulmonary hypertension. *Eur Respir J*. 2019;53(1):1801913.
 69. Kumar KR, Cowley MJ, Davis RL. Next-generation sequencing and emerging technologies [published online ahead of print May 16, 2019]. *Semin Thromb Hemost*. <https://doi.org/10.1055/s-0039-1688446>.
 70. Costa-Silva J, Domingues D, Lopes FM. RNA-Seq differential expression analysis: an extended review and a software tool. *PLoS One*. 2017;12(12):e0190152.
 71. Ghiassian SD, et al. Endophenotype network models: common core of complex diseases. *Sci Rep*. 2016;6:27414.
 72. Hovi P, et al. Glucose regulation in young adults with very low birth weight. *N Engl J Med*. 2007;356(20):2053–2063.
 73. Zanetti D, Tikkanen E, Gustafsson S, Priest JR, Burgess S, Ingelsson E. Birthweight, type 2 diabetes mellitus, and cardiovascular disease: addressing the barker hypothesis with mendelian randomization. *Circ Genom Precis Med*. 2018;11(6):e002054.
 74. Goss KN, et al. Early pulmonary vascular disease in young adults born preterm. *Am J Respir Crit Care Med*. 2018;198(12):1549–1558.
 75. Maron BA, Abman SH. Translational advances in

- the field of pulmonary hypertension. Focusing on developmental origins and disease inception for the prevention of pulmonary hypertension. *Am J Respir Crit Care Med.* 2017;195(3):292–301.
76. Tobi EW, et al. DNA methylation of IGF2, GNASAS, INSIGF and LEP and being born small for gestational age. *Epigenetics.* 2011;6(2):171–176.
77. Lewis GD, et al. Metabolic profiling of right ventricular-pulmonary vascular function reveals circulating biomarkers of pulmonary hypertension. *J Am Coll Cardiol.* 2016;67(2):174–189.
78. Kaati G, Bygren LO, Edvinsson S. Cardiovascular and diabetes mortality determined by nutrition during parents' and grandparents' slow growth period. *Eur J Hum Genet.* 2002;10(11):682–688.
79. Sharma A, et al. Integration of molecular interactome and targeted interaction analysis to identify a COPD disease network module. *Sci Rep.* 2018;8(1):14439.
80. Sutton TR, et al. A robust and versatile mass spectrometry platform for comprehensive assessment of the thiol redox metabolome. *Redox Biol.* 2018;16:359–380.
81. Otterbein LE, et al. Carbon monoxide has anti-inflammatory effects involving the mitogen-activated protein kinase pathway. *Nat Med.* 2000;6(4):422–428.
82. Wang Z, et al. Gut flora metabolism of phosphatidylcholine promotes cardiovascular disease. *Nature.* 2011;472(7341):57–63.
83. Karlsson FH, et al. Symptomatic atherosclerosis is associated with an altered gut metagenome. *Nat Commun.* 2012;3:1245.
84. Feng Q, et al. Integrated metabolomics and metagenomics analysis of plasma and urine identified microbial metabolites associated with coronary heart disease. *Sci Rep.* 2016;6:22525.
85. Rossi G, Ignatiadis M. Promises and pitfalls of using liquid biopsy for precision medicine. *Cancer Res.* 2019;79(11):2798–2804.
86. Francis R, Lewis C. Myocardial biopsy: techniques and indications. *Heart.* 2018;104(11):950–958.
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.* 2017;38(7):511–515.
88. Xie J, Yang J, Hu P. Correlations of circulating cell-free DNA with clinical manifestations in acute myocardial infarction. *Am J Med Sci.* 2018;356(2):121–129.
89. Wu PY, Cheng CW, Kaddi CD, Venugopalan J, Hoffman R, Wang MD. -Omic and electronic health record big data analytics for precision medicine. *IEEE Trans Biomed Eng.* 2017;64(2):263–273.
90. Rumsfeld JS, Joynt KE, Maddox TM. Big data analytics to improve cardiovascular care: promise and challenges. *Nat Rev Cardiol.* 2016;13(6):350–359.
91. Ouyang L, Apley DW, Mehrotra 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(e1):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 performance. *Philos Trans A Math Phys Eng Sci.* 2019;377(2142):20180144.
93. Cirillo D, Valencia A. Big data analytics for personalized medicine. *Curr Opin Biotechnol.* 2019;58:161–167.
94. Fogel AL, Kvedar JC. Artificial intelligence powers digital medicine. *NPJ Digit Med.* 2018;1:5.
95. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. *Nat Med.* 2019;25(1):44–56.
96. Sardar P, Abbott JD, Kundu A, Aronow HD, Granada JF, Giri J. Impact of artificial intelligence on interventional cardiology: from decision-making aid to advanced interventional procedure assistance. *JACC Cardiovasc Interv.* 2019;12(14):1293–1303.
97. Guyton AC, Coleman TG, Granger HJ. Circulation: overall regulation. *Annu Rev Physiol.* 1972;34:13–46.
98. Ong MS, et al. Learning a comorbidity-driven taxonomy of pediatric pulmonary hypertension. *Circ Res.* 2017;121(4):341–353.
99. Lee JH, et al. Phenotypic and genetic heterogeneity among subjects with mild airflow obstruction in COPDGene. *Respir Med.* 2014;108(10):1469–1480.
100. Maron BA, Cockrill BA, Waxman AB, Systrom DM. The invasive cardiopulmonary exercise test. *Circulation.* 2013;127(10):1157–1164.
101. Oldham WM, et al. Network analysis to risk stratify patients with exercise intolerance. *Circ Res.* 2018;122(6):864–876.
102. Antman E, Weiss S, Loscalzo J. Systems pharmacology, pharmacogenetics, and clinical trial design in network medicine. *Wiley Interdiscip Rev Syst Biol Med.* 2012;4(4):367–383.
103. Holmes MV, et al. Mendelian randomization of blood lipids for coronary heart disease. *Eur Heart J.* 2015;36(9):539–550.
104. Spiekerkoetter E, et al. Randomised placebo-controlled safety and tolerability trial of FK506 (tacrolimus) for pulmonary arterial hypertension. *Eur Respir J.* 2017;50(3):1602449.
105. Menche J, et al. Disease networks. Uncovering disease-disease relationships through the incomplete interactome. *Science.* 2015;347(6224):1257601.
106. Lee LY, Loscalzo J. Network medicine in pathobiology. *Am J Pathol.* 2019;189(7):1311–1326.
107. Maron BA, Opatowsky 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. *Eur J Heart Fail.* 2013;15(3):277–283.
108. Leopold JA, et al. Aldosterone impairs vascular reactivity by decreasing glucose-6-phosphate dehydrogenase activity. *Nat Med.* 2007;13(2):189–197.
109. Parikh VN, et al. MicroRNA-21 integrates pathogenic signaling to control pulmonary hypertension: results of a network bioinformatics approach. *Circulation.* 2012;125(12):1520–1532.
110. Bertero T, et al. The microRNA-130/301 family controls vasoconstriction in pulmonary hypertension. *J Biol Chem.* 2015;290(4):2069–2085.
111. Chen KH, et al. Epigenetic dysregulation of the dynamin-related protein 1 binding partners MiD49 and MiD51 increases mitotic mitochondrial fission and promotes pulmonary arterial hypertension: mechanistic and therapeutic implications. *Circulation.* 2018;138(3):287–304.
112. Sweatt AJ, et al. Discovery of distinct immune phenotypes using machine learning in pulmonary arterial hypertension. *Circ Res.* 2019;124(6):904–919.
113. Hemnes AR, et al. PVDOMICS: a multi-center 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.