Advances in Genetics

Systems-Based Approaches to Cardiovascular Biomarker Discovery

Francisco J. Azuaje, PhD; Frederick E. Dewey, MD; Dirk L. Brutsaert, MD, PhD; Yvan Devaux, PhD; Euan A. Ashley, MD, PhD; Daniel R. Wagner, MD, PhD

The pace of discovery of cardiovascular biomarkers seems to be slowing. Although important advances in the identification of molecular biomarkers have been made during the past decade, it is becoming apparent that their clinical relevance is limited and that advanced discovery methods are needed.

There are >27,000 articles on cardiac biomarkers in PubMed. However, only a small number of these biomarkers are in clinical use. In patients with heart failure (HF), numerous biomarkers have been evaluated (Table), but only brain natriuretic peptide and its precursor molecules have been widely applied. In patients with stable coronary artery disease, C-reactive protein (CRP) is a potential prognostic biomarker. Although there is conflicting evidence about the ability of CRP to augment the predictive accuracy of traditional clinical risk factors, targeting patients with elevated CRP for lipid lowering therapy with rosuvastatin was associated with improved cardiovascular outcomes in 1 randomized controlled trial. However, a recent meta-analysis of 83 studies including >60,000 patients has questioned its value. The key hurdle to the introduction of new “omic” biomarkers has been the inability to demonstrate their clinical relevance and validity. This refers not only to meeting analytic validity and independent evaluation criteria, but also to the biomarkers’ capacity to guide the improvement of patient outcomes. We believe that a more integrated, unbiased approach is essential to discover novel diagnostic and prognostic models of heart disease.

Biomarker discovery research has traditionally emphasized the study of individual molecular indicators of clinical condition. A major limitation of this approach is that the informational complexity underpinning many clinical states are not adequately taken into account.

The relevance of hypothesis-driven, reductionist approaches to biomarker discovery is indisputable. However, an overreliance on this strategy may limit the translation of fundamental research into new clinical applications. Limitations to traditional research should be seen as logical consequences of the limited ability to interrogate the multivariate and combinatorial characteristics of cellular networks implicated in multi-factorial common diseases. Biological networks sense, integrate and compute responses to thousands of signals in an integrative fashion. Moreover, the networks controlling physiological responses are dynamic, cell-specific and organized into complex functional hierarchies.

Systems approaches can guide the identification of single biomarkers based on a deeper understanding of their underlying biology. In addition, they can enable strategies in which multiple biomarkers are assayed together, as a system, either in blood or tissue. In this context, biomarkers become clinical state indicators computationally inferred from networks of interacting molecular entities. Thus, biomarkers can represent measurements of network activity comprising not only the expression of biochemical substances (or their combinations), but also information on clinically meaningful biological interactions.

In cardiovascular research, the potential of systems-based methodologies is being demonstrated in several application domains. They include atherosclerosis, HF, and other cardiovascular pathology measured in either tissue or blood. Such biomarkers have been implicated in biological processes as disparate as angiogenesis, inflammation, and oxidative stress. The discovery of such widely applicable biomarkers has been facilitated by advances in high-throughput molecular profiling, including transcriptomics and proteomics.

In this review we discuss how the integration of data arising from different molecular profiling techniques are enabling a more detailed and biologically meaningful characterization of putative biomarkers. Such an integration of technologies and disciplines promises a more accurate classification of clinical conditions, while presenting opportunities for novel therapeutic interventions.

Why Does Systems Research Matter in Cardiovascular Biomarker Discovery?

The accumulation of relatively large volumes of “omic” data has not yet radically altered the way cardiovascular diseases are detected or treated. Nevertheless, it is becoming increasingly accepted that our understanding of physiological states can be illuminated through the application of information processing principles. This goes beyond the mere application of statistical
The application of systems approaches to biology and physiology goes back to times preceding the molecular biology revolution. However, pregenomic era systems research was naturally hindered by the absence of the technologies necessary to accomplish such an integration, and by our ability to comprehensively interrogate such systems. Significant reductions in the cost of data generation at different resolution are making systems-based research feasible across application domains.

Systems approaches to the study of complex phenotypes, such as human cardiovascular disease, may more accurately reflect the underlying mechanisms than traditional reductionist approaches. Complex cardiovascular diseases can be viewed as the end result of multi-scale interactions between individual molecular-, cellular-, organ-, and organism-level entities and environmental factors. In the lexicon of network-based modeling of systems biology, we can thus characterize these complex phenotypes as a complex behavior, or emergent property, (i.e., collectively driven by multi-scale biological interaction networks). Systems approaches seek a more holistic understanding of these interactions and the resultant disease phenotypes.

Integration of systems-based approaches and traditional reductionist approaches to biomarker discovery is thus a new opportunity in cardiovascular research. In HF research, for example, there is already recognition of the capacity of systems approaches to elucidate clinical responses based on the integration of multiple biological information sources.

The translation of traditional biomarker research into clinically relevant predictors of disease and treatment is limited by several factors. One is that single-gene biomarkers cannot fully explain complex phenotypes. Statistically speaking, it is unlikely that individual measurable biological characteristics will explain a significant proportion of phenotypic variation. Furthermore, gene activity variability within and across patients is expected to be high because of inherent biological noise, as well as the complex interplay of sensors and drivers of disease.

Systems-based approaches may identify putative biomarkers without necessarily expecting them to be (statistically) significantly differentially expressed across clinical categories. Potential biomarkers can in principle be weakly differentially expressed without compromising their integrated capacity to classify clinical states while helping to elucidate underlying biology. Genes with significant individual discriminatory capacity tend to represent “passenger” genes, whereas “driving” disease networks can include combinations of relatively weakly differentiating genes. Nevertheless, this connection has not been conclusively shown in other “omic” information types.

Thus, the concept of biomarker, when viewed in a systems context, is significantly enriched. Biomarkers are clinical-state indicators that can be derived from the integration of multiple network-driven biological activity levels. Such synergistic molecular activity information can be estimated by mapping patient-derived profiling data onto clinically meaningful interaction maps. In addition, the resulting integrated activity levels can be associated with specific biological pathways, protein complexes and other functional modules. Figure 1 summarizes important differences between traditional and systems-based biomarker research.

### Interactions, Networks, and Modules

A network represents a logical approach to characterizing multidimensional complex interactions. As such, network representations are suited to the description of disease-associated molecular states. Networks can be graphically represented as a collection of nodes and edges. Nodes define different types of molecular entities (eg, genes) and edges describe relationships between nodes, such as gene coexpression. Figure 2 illustrates an example of a protein-protein interaction (PPI) network implicated in inflammation and myocardial infarction (MI). Proteins and their physical interactions are shown as dots and lines respectively.

Nodes and edges can be characterized with mathematical descriptors that are biologically meaningful. For instance, the number of edges associated with a node is the degree of the node, and the number of shortest paths connecting any 2 nodes in the network defines the betweenness centrality of the node. High-degree nodes are commonly referred to as hubs. Nodes with high centrality are often labeled as bottleneck nodes. Nodes displaying either (or both) properties have been shown to be potential biomarkers.

Networks can be inferred as global representations of underlying molecular profiles that are common to different physiological states or phenotypes. Also it is possible to define phenotype-specific networks, for example by describing relationships observed in patients within a prognosis category. Such dynamic network topology is likely a better reflection of underlying biology than static network representations.

Subnetworks of highly connected nodes, that is, groups of nodes that are in close graphical or biological functional proximity, can be identified. This can lead to the determination of network modules. Modularity has been suggested as a fundamental property of many complex systems. Modules are detected with computational tools and can represent strong indicators of clinical states. For example, genes grouped in a module can together exhibit differential expression across patient groups, or can be significantly implicated in clinically related molecular perturbations. Figure 2B shows a partial view of a module identified in an inflammation-related PPI network.

Figure 3 illustrates a typical biomarker discovery framework. Information extracted from different resources is

<table>
<thead>
<tr>
<th>Table. Examples of Biomarkers in Heart Failure</th>
</tr>
</thead>
<tbody>
<tr>
<td>CRP (C-reactive protein)</td>
</tr>
<tr>
<td>TNF (tumor necrosis factor α)</td>
</tr>
<tr>
<td>MMP2 (matrix metalloproteinase-2)</td>
</tr>
<tr>
<td>MMP9 (matrix metalloproteinase-9)</td>
</tr>
<tr>
<td>Oxidized low-density lipoproteins</td>
</tr>
</tbody>
</table>

Downloaded from http://circgenetics.ahajournals.org/ by guest on October 30, 2017
processed to generate interaction networks. The structures of the resulting networks, as well as their associations with clinical states, are analyzed with different computational techniques. This also involves the integration of additional information, such as molecular biological function (“gene ontologies”). This can be mapped onto the networks or their individual components. The typical outcomes of this phase are sets of functionally related components (e.g., nodes) and associations between them and clinical states. This information can reduce the search space for potential biomarkers. Different selection procedures can be implemented according to the application investigated and the resources available for independent validations. Candidate biomarkers may then be applied as inputs to computational models of clinical classification, followed by multiple independent evaluations.

**Networks for Biomarker Discovery: Where Do They Come From?**

Network interactions are typically obtained from either published literature, including those annotated in interaction databases or the application of computing tools to different types of omics data. Genetic, protein-protein, protein-DNA, and metabolic interactions are commonly found in the first category. Many of the network inference algorithms applied to cardiovascular biomarker discovery are based on the detection of gene relationships observed in single-source datasets, such as gene coexpression data. However, with the

---

**Figure 1.** Comparison of traditional and systems-based biomarker research. System-based approaches facilitate the translation of discoveries to novel clinical applications.

**Figure 2.** A protein interaction network in cardiovascular biomarker research. **A**, Global view of a PPI network implicated in inflammation and MI. Adapted from Azuaje et al under Creative Common License.14 **B**, Partial view of a network module. Dots and lines represent proteins and their physical interactions respectively. Modules can be associated with clinically relevant processes and may include candidate biomarkers.
increasing availability of high-throughput data, these algorithms are being adapted to multiple data sources. For an overview of tools to generate and analyze biological interaction networks, the reader is referred to Ghosh et al.17

Another approach to network information generation consists of overlaying different types of data on networks extracted from the literature or inferred from data.18 In this case, the network may be seen as a discovery “scaffold” that supports the identification of novel insights across omic data types or clinical categories. Using PPI networks and plasma-derived proteomics data, for example, Jin et al predicted major adverse cardiac events.19 Candidate biomarkers were represented by sets of interacting proteins with differential protein concentration levels in case-control groups.

Key Strategies for Systems-Based Biomarker Discovery in Cardiovascular Research
Recent advances can be assigned to 2 discovery strategies (Figure 4).14,20 In the first category, candidate biomarkers are identified after establishing associations between network structure properties and clinical phenotypes. In the second category, computing programs are implemented to search for subnetworks as putative biomarkers. Figure 5 depicts examples of advances in network-based cardiovascular biomarkers.

Discovery Strategy 1 (DS1): Linking Network Structure to Clinical States
Investigation of network topological properties has laid down significant connections between clinical states and their underlying biology in cardiovascular research (Figure 4A to C). We and others have demonstrated that hubs and bottlenecks can represent potential novel biomarkers in HF and ventricular remodeling.14,21,22 For instance, TRAF2 (TNF receptor-associated factor 2) and ubiquitin C are bottlenecks in a PPI network related to MI and inflammation that exhibit prognostic potential.14 Although the diagnostic specificity of these genes requires further assessment, such findings contribute to the understanding of disease biology. These associations can be identified, for instance, by detecting strong correlations between node degree and gene expression from patients across clinical categories. Based on the analysis of a coexpression network inferred from endothelial cell gene expression data, Romanoski et al showed that hubs HMOX1 (heme oxygenase 1) and CHAC1 (cation transport regulator homolog 1) represent candidate biomarkers of oxidized phospholipid response.23

Another approach comprises the identification of modules with prominent functional roles. This is commonly achieved by testing the hypothesis that the candidate module is significantly functionally compact, (eg, module genes are statistically associated with specific biological processes). Thus, relevant modules may offer the basis for candidate biomarkers, including their individual members or their combinations.

Dewey et al extracted gene coexpression network modules, which were shown to distinguish cardiac hypertrophy and failure.24 They identified a novel transcription factor, ZIC2 (transcription factor zinc finger protein 2) associated with modules common to developing and failing myocardium. In a cohort of dilated cardiomyopathy and control samples, Lin et al established dilated cardiomyopathy–specific network modules based on the combination of information from a PPI network and tissue gene expression data.21 They showed how the expression levels of the genes in these modules can effectively classify patients. Civelek et al detected modules in a gene coexpression network generated from endothelial cell samples with different levels of atherosclerosis susceptibility.25 Their analysis resulted in a signature of endoplasmic reticulum stress that estimates atherosclerosis propensity.
Using mass spectrometry data from tissue sections, Stegmann et al built lipid coexpression networks to investigate novel atherosclerosis-specific associations. In these networks, lipids and lipid-lipid concentration correlations represented nodes and edges respectively. They found lipid clusters that can distinguish stable and unstable lesions.

**Discovery Strategy 2 (DS2): Identification of Subnetworks With Differential Molecular Responses**

Approaches included in DS1 may be defined as unsupervised approaches. This is because the detection of critical nodes and modules does not necessarily depend on the incorporation of information about clinical phenotypes. In contrast to DS1, DS2 mostly represent supervised discovery approaches. In this case, the objective is to search for subnetworks that meet different quality criteria. This may include network structure constraints and the capacity of these subnetworks to distinguish between clinical states.

Each subnetwork, as an integrated unit, can be defined as a new biomarker. In this case, classification capacity is estimated based on the combined molecular activity of candidate subnetworks. Thus, biomarkers are encoded as network activity scores. This can be performed by computing different measures, such as those based on the mean expression values observed in the subnetwork (Figure 4D).

In cardiovascular research, 2 main types of approaches have been reported so far. The first searches for differentially expressed gene sets among lists of user-selected pathways. Such gene sets typically originate from curated repositories of molecular pathways. Azuaje et al searched for candidate biomarkers of ventricular dysfunction after MI among hundreds of pathways. Blood-derived gene expression data were laid out on patient-specific gene sets to detect those differentially expressed across patient groups.

A second category aims to identify predictive subnetworks, which are extracted from networks describing relationships among genes in specific clinical states (eg, gene coexpression networks associated with good and poor prognosis). Azuaje et al and Nepomuceno et al defined post-MI prognosis-related gene networks. These networks were based on blood-derived gene expression correlation coefficients and trees of mathematical models respectively. Azuaje et al reported
differentially activated subnetworks of small sizes, whose mean and pair-wise differences in expression values accurately classified patients in an independent dataset. The top subnetwork was defined by angiogenesis-implicated genes: vascular endothelial growth factor B (VEGFB), death-domain associated protein (DAXX), tenascin XB (TNXB), and latent transforming growth factor beta binding protein 4 (LTBP4). Nepomuceno et al generated partially overlapping networks for good and poor prognosis as defined by the patient’s ejection fraction. Their networks assign patients to prognostic categories by matching the patients to class-specific networks. Patient-network matching is determined according to the capacity of the underlying mathematical models to estimate the patient’s gene expression profile. Thus, a patient is assigned to the class whose corresponding network best fits the patient’s expression signature. An attractive quality of such approaches is that the procedure by which molecular interactions are defined is not constrained to known pathways, allowing for the identification of novel biomarker pathways that may illuminate new disease biology.

Clinical Challenges and Perspectives
The translational potential of network-derived biomarkers may be hampered by the incomplete nature of the networks investigated. They represent approximate views of the complete set of factors and relationships underlying disease states. Nevertheless, higher quality will increasingly be accomplished through the availability of larger datasets associated with carefully characterized clinical phenotypes. In addition, more cost-effective ways to integrate multiple information views will be available. This is also important to address the challenge of linking tissue-derived biomarkers to their circulation or to novel counterparts in the blood. The latter constitutes an important requirement to demonstrate the clinical applicability of network-derived biomarkers.

Systems approaches are useful to identify new biomarkers with potential phenotype-driving roles, such as microRNAs. We recently illustrated the potential of network analysis to identify candidate biomarkers among the microRNA family. Inflammation network modules with significant involvement in transcriptional regulation were shown to be
targeted by a small set of microRNAs. Other studies have also demonstrated the usefulness of microRNA network approaches. Concomitant with further identification of novel microRNA families, systems biology of microRNAs will enable novel biomarker discovery.

Challenges arise in relation to the assessment of result concordance across network-based studies. This will not only require the standardization of data acquisition protocols, but also of information exchange and computational validation tasks. Advances will also depend on our capacity to share data and software. The former is currently facilitated by large public repositories of high-throughput molecular data, such as the Gene Expression Omnibus.

The technical reproducibility of microarray measurements has been established and the robustness of disease signatures across independent datasets has been demonstrated. The latter, for instance, was illustrated by Dudley et al in thousands of public datasets and hundreds of disease states. Challenges in the standardization of different types of biomarker assays are also being addressed through quality assurance procedures that are suitable to routine biomarkers. The analytic robustness of applications based on whole-genome sequencing technologies have been investigated, and recommendations for achieving accurate and concordant results across different platforms have been made. These investigations further highlight the potential to improve assay robustness and predictive reproducibility of systems-based biomarker discovery.

Systems approaches to biomarker discovery can improve predictive capability in comparison with traditional biomarker research. Researchers have also shown that systems approaches can enhance prediction reproducibility across independently generated datasets. Network-derived biomarkers can successfully address this demand in different ways. First, this approach reduces dependence on analysis techniques that are unsuitable to datasets typically characterized by small numbers of samples and large numbers of highly-correlated genes. Second, the search for biomarkers can be guided by prior knowledge of canonical pathways and simultaneously explore previously uncharacterized physiological pathways that, when perturbed, may produce disease phenotypes. As such, network-based approaches to biomarker discovery can illuminate novel disease biology and new therapeutic targets in addition to providing clinically-relevant diagnostic and prognostic markers. New clinical applications of systems approaches will be facilitated by increasing access to larger biological knowledge resources.

Network-derived biomarkers can represent “drivers” of diseases, which are expected to exhibit more robust behavior than “passenger” entities. Moreover, as the focus of discovery moves from gene lists to networks, the number of potentially spurious associations and variability are reduced. This is possible because the search space of potential biomarkers is reduced, and gene-gene relationships are explicitly modeled. In the near future, systems approaches will provide opportunities to identify novel therapeutic targets based on a deeper understanding of pathogenetic mechanisms.

Notwithstanding the above-mentioned advances and opportunities, emerging systems-based biomarkers await to be validated in the routine clinical setting. This will require reproducing the predictive capability of the most promising models in independent prospective studies.

A systems-based thinking is required to take advantage of these opportunities. Stronger integration between the clinical, computing, and biological sciences will be essential for the successful translation of network-derived biomarkers into the clinical setting. As part of this challenge, there is a need to integrate network-based biomarkers with those derived from imaging and electrophysiological signals. To move beyond research possibility into clinical reality, advances will also entail major changes in the way patient records are stored and interpreted. This will necessitate a new type of cardiologist with a stronger proficiency in systems science.

Conclusion

The integrative analysis of biological networks provides a framework for predictive and personalized medicine. The traditional notion of biomarkers is being reformulated. This is giving way to the analysis of biological networks to prevent, identify, and treat disease. The clinical acceptance of systems approaches will depend on the integration of these network views across different biological organization levels from subcellular to patient population and back. Systems-based and traditional reductionist approaches complement each other and enhance potential clinical impact. Closer cooperation between clinicians, epidemiologists, biologists, and computing scientists will be essential to advance cardiovascular systems biology. To succeed in this challenge, cardiologists will develop new expertise to derive clinically meaningful knowledge from complex biological interaction networks. Independent validations based on prospective clinical studies are required.
SOURCES OF FUNDING

Authors based in Luxembourg acknowledge support from FNR, the Society for Research on Cardiovacular Diseases and MCESR. F.E.D. was supported by NIH/NHLBI training grant T32 HL04274-01A2 and the Stanford University School of Medicine Dean’s Postdoctoral Fellowship. Dr Ashley was supported by NIH DP2 OD004613 and NIH R01 HL105993.

DISCLOSURES

None.

REFERENCES

Systems-Based Approaches to Cardiovascular Biomarker Discovery
Francisco J. Ázuaje, Frederick E. Dewey, Dirk L. Brutsaert, Yvan Devaux, Euan A. Ashley and
Daniel R. Wagner

Circ Cardiovasc Genet. 2012;5:360-367
doi: 10.1161/CIRCGENETICS.112.962977
Circulation: Cardiovascular Genetics is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231
Copyright © 2012 American Heart Association, Inc. All rights reserved.
Print ISSN: 1942-325X. Online ISSN: 1942-3268

The online version of this article, along with updated information and services, is located on the World Wide Web at:
http://circgenetics.ahajournals.org/content/5/3/360

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

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

Subscriptions: Information about subscribing to Circulation: Cardiovascular Genetics is online at:
http://circgenetics.ahajournals.org/subscriptions/