Robust Physiological Metrics From Sparsely Sampled Networks

Alan A. Cohen[1,2,3]*, Sebastien Leblanc[4] and Xavier Roucou[4]

[1] Groupe de Recherche PRIMUS, Département de Médecine de Famille et de Médecine d'Urgence, Université de Sherbrooke, Sherbrooke, QC, Canada, [2] Centre de Recherche, Centre Hospitalier Universitaire de Sherbrooke (CRCHUS), Sherbrooke, QC, Canada, [3] Research Center on Aging, CIUSSS-de-l'Estrie-CHUS, Sherbrooke, QC, Canada, [4] Département de Biochimie et de Génomique Fonctionnelle, Université de Sherbrooke, Sherbrooke, QC, Canada

Physiological and biochemical networks are highly complex, involving thousands of nodes as well as a hierarchical structure. True network structure is also rarely known. This presents major challenges for applying classical network theory to these networks. However, complex systems generally share the property of having a diffuse or distributed signal. Accordingly, we should predict that system state can be robustly estimated with sparse sampling, and with limited knowledge of true network structure. In this review, we summarize recent findings from several methodologies to estimate system state via a limited sample of biomarkers, notably Mahalanobis distance, principal components analysis, and cluster analysis. While statistically simple, these methods allow novel characterizations of system state when applied judiciously. Broadly, system state can often be estimated even from random samples of biomarkers. Furthermore, appropriate methods can detect emergent underlying physiological structure from this sparse data. We propose that approaches such as these are a powerful tool to understand physiology, and could lead to a new understanding and mapping of the functional implications of biological variation.

Keywords: network physiology, complex dynamic system, big data, alternative proteins, systems biology, statistical distance

INTRODUCTION

Complex systems theorists have long viewed biological networks as one of the prime examples of complex systems (Holland, 1992; Mobus and Kalton, 2014), but biologists themselves, with some notable exceptions (e.g., Kitano, 2002; Tieri et al., 2010), have often been more reticent, preferring to understand biological signaling pathways from a more linear and reductionist perspective. This is starting to change, with systems biology gradually moving from simple inventories of large numbers of molecules to network-based approaches [e.g., Ingenuity Pathway Analysis, Kyoto Encyclopedia of Genes and Genomes (KEGG)-pathways, gene ontology (GO)-terms (Kanehisa and Goto, 2000; Harris et al., 2004; Krämer et al., 2014)]. Indeed, this entire Research Topic, one of the largest in the history of Frontiers Research Topics, is devoted to Network Physiology.

Indeed, drawing on insights from classical network theory (Albert and Barabási, 2002), over the last ~10 years, network physiology, and the related field of network medicine, have been making great strides in bringing network thinking into the biomedical realm [Network medicine tends to focus more on genetic and molecular networks (Barabási et al., 2011), while network physiology

focuses more on the temporal coordination of physiological function across systems, often at higher levels of organization, such as organs and organ systems (Ivanov et al., 2016), though the distinction is not always clear]. These methods are contributing numerous insights, such as coordination of sleep cycles across brain waves and muscle groups (Bashan et al., 2012) or understanding motor neuron function in *Caenorhabditis elegans* (Yan et al., 2017). Perhaps one of the most important insights is simply the repeated validation that there is coordination across biological networks, and thus that networks must be studied as ensembles, not as pieces. For example, there is now clear evidence that temporal dynamics of organ function show clear coordination across organ systems (Bartsch et al., 2015). Accompanying such insights come methods to quantify these network dynamics, such as time delay stability (TDS) and accompanying graphical methods (Bashan et al., 2012; Bartsch et al., 2015). It doesn't require much thinking about the principles of evolution, optimization, and organismal function to understand why this is expected; the surprise is that physiology existed for so long without looking for this coordination, or without considering the higher-order perspective that it implies. For the purposes of this review, an additional key insight is that this coordination leads to a limited number of largely discrete states of the larger system, with multiple subsystems changing together in abrupt and coordinated fashion (Bashan et al., 2012; Bartsch et al., 2014, 2015). For example, multiple organ systems change together at transitions between sleep stages (Bartsch et al., 2015). Similar principles apply at other hierarchical levels of biology, for example in β-cell regulation in response to glucose (Podobnik et al., 2020). This leads to the important conclusion that there are *biological attractor states*, discrete states toward which physiology/biology converges, and between which it shifts.

Nonetheless, there are still major challenges in applying classical network theory to many aspects of biological networks. Classical network theory is based on the ability to map networks relatively exhaustively in order to estimate properties such as connectedness, modularity, etc. (Strogatz, 2001; Barabási, 2016), but biological networks imply multiple levels of organization, interactions of different types of structure (physical, informational, etc.), and networks that, on the molecular level, are still very poorly mapped. Additional approaches are thus needed for when our sparse understanding of network structure limits the applications of classical approaches. Here, we briefly discuss the structure of organisms from a complex systems perspective and how much we do and don't know about their underlying networks. There are obviously a daunting variety of organisms and many levels of organization within them; we try to stay at a general level to enunciate principles that will apply to studies of any biological systems at the organism level or lower that are composed of networks, whether they be biochemical networks within cells or networks of tissues or brain regions. We then argue that, despite an imperfect knowledge of the finest-scale details, there is a coherence to biological states that suggests that we should be able to measure organismal state even with highly imperfect knowledge of the underlying networks. We use this framework to review a number of methods to infer organismal state via sparse sampling of networks, and to suggest

future avenues for further development of such methods. Our previous research in aging leads us to use examples primarily from this field, but the conclusions are much more general.

AN ORGANISM AS A COMPLEX DYNAMIC SYSTEM

There can be little doubt that organisms are complex dynamic systems, as they exhibit all the hallmarks of such systems: they are composed of multiple elements which interact with feedback (and also feedforward) mechanisms; the elements are organized both hierarchically and modularly; and there are clear emergent properties at the various hierarchical levels which emerge from the underlying dynamics at adjacent levels (e.g., Podobnik et al., 2017, 2020). These properties are also shared by other key examples of complex dynamic systems: economies, ecosystems, weather systems, societies, traffic systems, etc. However, organisms—and, more broadly, sub-organismal biological systems—exhibit some unique features that are not shared by all complex dynamic systems:

1. **They are highly optimized via natural selection** (Kriete, 2013). While systems such as economies or ecosystems may undergo weak selective pressure that affects their evolution, biological systems have had billions of years to become fine-tuned, with progress made in each generation conserved and transmitted via the genetic code. Molecular signaling pathways, for example, are much more fine-tuned than economic regulatory policy.

2. **Accordingly, they are goal-directed**. Weather systems, ecosystems, and most other complex dynamic systems simply exist, without an effort to achieve anything. Biological systems, however, have been fine-tuned for a reason: to maximize organismal fitness (roughly speaking). And in turn, for most of the biological machinery, this equates to optimizing physiological/biological equilibrium in the face of constant internal and environmental variation (Cohen et al., 2012; Podobnik et al., 2017, 2020), as denoted by concepts such as homeostasis, homeodynamics, robustness, and resilience (Ives, 1995; Sterling, 2015; Ukraintseva et al., 2016). In other words, biological systems are *designed* (though not in a theological sense) to maintain organismal balance.

3. **Biologically relevant information is conserved over long timescales**. The genetic code permits the conservation of information across very long timescales—billions of years. Many other complex dynamic systems do not really conserve information at all, and those that do are either much more recent (e.g., writing in economic systems, code in computer systems), and/or the information is much less precisely transmitted (cultural transmission).

4. **Accordingly, biological systems are perhaps the most complex systems that exist**. The combination of natural selection, very long timescales, goal-directedness, and information conservation has permitted biological systems to become what is likely the most structured complex

system known. New levels and layers of biological organization are constantly being discovered (see below).

It is this last point, perhaps, that explains why biologists are so far behind other fields in adapting complex systems thinking. There was simply too much information, and too much structure, for classic complex systems methodologies to be broadly useful. The advent of -omics and imaging technologies and big-data capabilities are changing this now, but the challenge remains that our basic knowledge of biological organization is still rudimentary.

WHAT WE DO AND DON'T KNOW ABOUT BIOLOGICAL ORGANIZATION
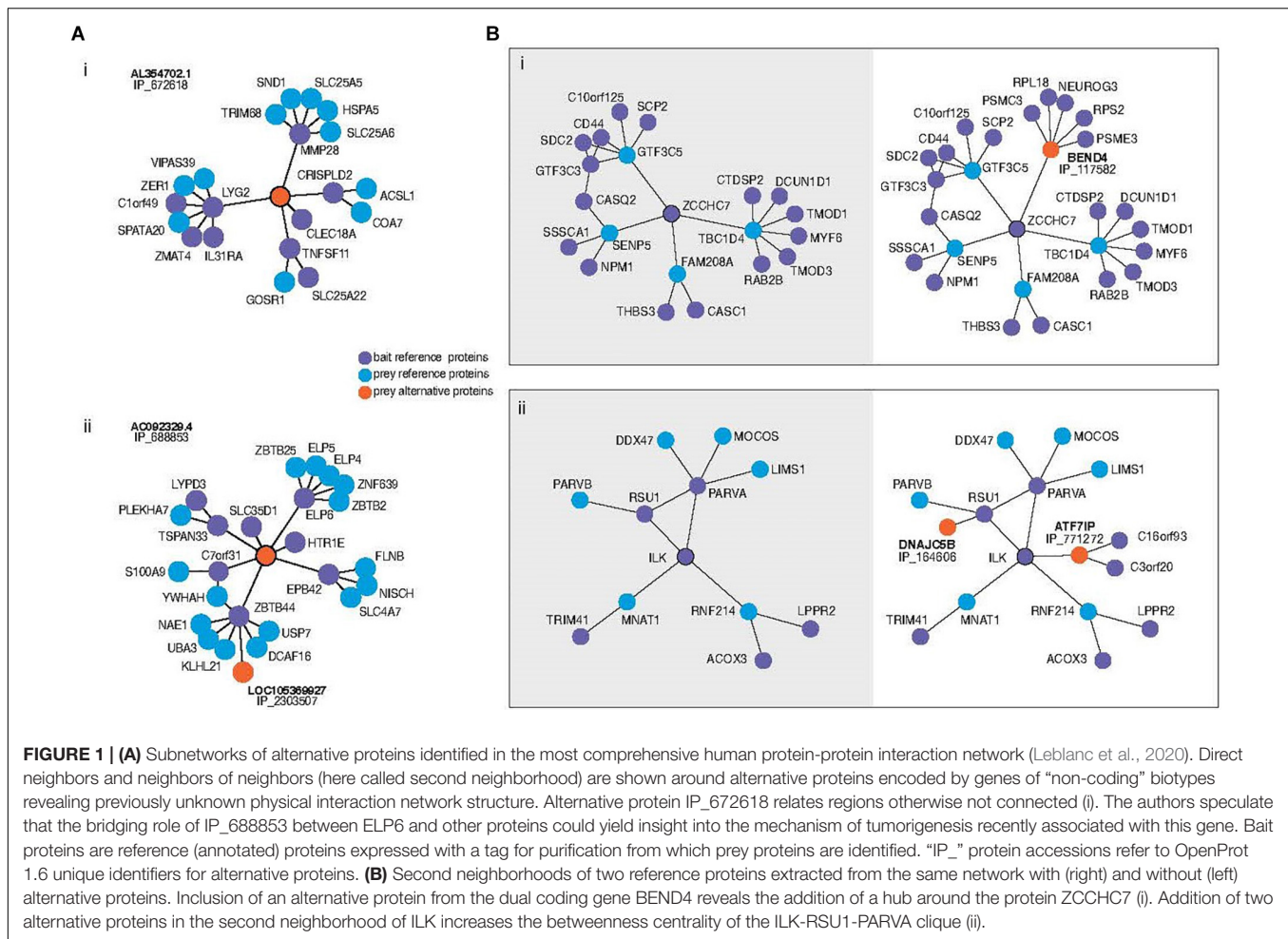
At least since the discovery of the structure of DNA, our knowledge of biological organization has been increasing rapidly. As this process has unfolded, there has been a repeated theme: scientists at each stage fail to appreciate how much more detail has yet to be uncovered. For example, new hormones—key molecules for regulatory coordination—are regularly discovered, such as hepcidin (Nemeth et al., 2003) and apelin (Vinel et al., 2018). Perhaps the best example is the changes that have occurred to biochemistry's "central dogma" over the years. Once it was established that DNA is transcribed into messenger RNA (mRNA), and that mRNA is translated into proteins, this transfer of information (genes—> mRNA—> protein) became known as the "central dogma" (Crick, 1970). Proteins were considered the key biological effectors. However, it was not long before this dogma began to break down at a number of levels: reverse transcriptases from viruses such as HIV can reverse the flow of information (Spiegelman et al., 1971). Perhaps more importantly, the central dogma radically underestimates the sophistication of the information processing. A single gene can be spliced into multiple mRNAs (Chow et al., 1979). Many new types of RNAs, such as micro RNAs and snoRNAs, are being discovered, often with key signaling roles (Scott and Ono, 2011).

Here, we will look in detail at the recent discovery of "alternative proteins" (Brunet et al., 2020), a telling example of how new layers of biological organization are constantly being uncovered, and how each time they are, our understanding as relates to the application of complex systems methods to biology would need to be re-thought. Since the discovery of alternative mRNA splicing in eukaryotes– i.e., that a single gene could produce many mRNA isoforms—part of the revised central dogma, implicit or explicit, has been that a mature mRNA (i.e., one that has already undergone all splicing and processing before being translated into a protein) codes for one and only one protein. A few exceptions had been detected, but they were considered just that: exceptions. However, with the advent of high-throughput technologies, it was recognized in the early 2010s that most mature mRNAs contained multiple start codons, the series of three nucleotides in the mRNA that signals the ribosome to start translation (Ingolia et al., 2011; Vanderperre et al., 2013). As a result, it was possible that multiple proteins were being translated from a single eukaryotic mRNA. By combining

the presence of start and stop codons in the mRNA sequences, it was predicted that the average mature mRNA might code for 7.8 proteins, with some mature mRNAs coding for as many as 89 (Open Prot database v1.6: Brunet et al., 2019), leading to the possibility that the proteome was many-fold larger than previously thought. Indeed, subsequent experiments have shown that many of the so-called "alternative proteins"—proteins that are coded by a mature mRNA, other than the "reference protein" that was canonically expected—are indeed produced by cells. They have been detected by mass spectrometry and by ribosomal profiling, are often highly conserved evolutionarily, and in some cases they have been shown to have crucial roles in biological processes and diseases, often working in tandem with other proteins produced by the same mRNA (Samandi et al., 2017; Chen et al., 2020). Hence, in contrast to the general belief that polycistronic mRNAs are restricted to prokaryotes, an ever-increasing fraction of mRNAs in eukaryotes are known to encode at least two different proteins.

Not all alternative proteins predicted by mRNA sequences have been shown to be translated or to have important biological roles, and there is currently substantial uncertainty as to how many of the alternative proteins will prove biologically important, but the number would certainly appear to be in the thousands, if not much higher. Indeed, the line between "biologically important" and "translational noise" may not be so clear: many alternative proteins likely arise via mutations generating random start codons, but the proteins so-generated may have some biological activity. If that activity is strongly harmful, natural selection should quickly eliminate the mutation, but if the novel protein, due to biochemical stochasticity, is anywhere close to neutral, there may exist a long period where it is produced in small quantities and subject to gradual selection toward a beneficial role, even if it does not play a core regulatory function. There is thus substantial potential for alternative proteins to play broad but hard-to-pinpoint regulatory roles, even beyond the substantial number that are in the process of being identified as key players.

Among those alternative proteins that are emerging as having clear functions, recent work is permitting us to understand how they integrate into biological networks. Briefly, data generated by high throughput mass spectrometry (MS) experiments have been re-analyzed with the inclusion of alternative proteins in the library used at the spectra-peptides matching step (Leblanc et al., 2020). When these data result from separate purification of multiple tagged proteins followed by MS analysis it is possible to confidently identify alternative proteins in a network of protein physical interactions. Interestingly, a surprising number appear to have roles linking what otherwise appear to be separate modules, or as hubs (**Figure 1**). In other words, the structure of the regulatory network, including important higher-order network properties, could completely change with or without the inclusion of the alternative proteins. Indeed, the addition of alternative protein IP_688845 in the interactome of the ELP6 protein may yield insight surrounding its recent association with tumorigenesis and migration of melanoma cells (Close et al., 2012; **Figure 1Aii**). The re-analysis shows that the novel interactor bridges this member of the elongation complex

FIGURE 1 | (A) Subnetworks of alternative proteins identified in the most comprehensive human protein-protein interaction network (Leblanc et al., 2020). Direct neighbors and neighbors of neighbors (here called second neighborhood) are shown around alternative proteins encoded by genes of "non-coding" biotypes revealing previously unknown physical interaction network structure. Alternative protein IP_672618 relates regions otherwise not connected (i). The authors speculate that the bridging role of IP_688853 between ELP6 and other proteins could yield insight into the mechanism of tumorigenesis recently associated with this gene. Bait proteins are reference (annotated) proteins expressed with a tag for purification from which prey proteins are identified. "IP_" protein accessions refer to OpenProt 1.6 unique identifiers for alternative proteins. (B) Second neighborhoods of two reference proteins extracted from the same network with (right) and without (left) alternative proteins. Inclusion of an alternative protein from the dual coding gene BEND4 reveals the addition of a hub around the protein ZCCHC7 (i). Addition of two alternative proteins in the second neighborhood of ILK increases the betweenness centrality of the ILK-RSU1-PARVA clique (ii).

to protein clusters which include S100A9, recently identified as biomarker up-regulated in metastatic melanomas (Wagner et al., 2019) and TSPAN33, a protein which modulates cell adhesion and migration through its effect on plasma membrane mechanical properties (Navarro-Hernandez et al., 2020). Several hypotheses could be explored with this larger subnetwork as a starting point. While it is not yet proven whether these changes in our understanding of network structure are expected to lead to changes in the predicted functional dynamics, there is every reason to expect that this would be the case. In similarly generated data, alternative proteins have even been observed in the interaction network of viral proteins in Zika virus infected human cells (Leblanc and Brunet, 2020).

The changes in biological paradigm implied by the existence of a broad range of functional or semi-functional alternative proteins should have been expected. Why would occasional mutations not have produced additional start codons? Why would those additional proteins not have been translated? And why would natural selection not have then subsequently acted on them, leading in all likelihood to the sharing of a mature mRNA as a way to improve coding efficiency of the genome and coordination of related regulatory functions? We would argue that this example is a clear microcosm of how the combination

of stochasticity, biological information conservation, and billions of years of natural selection has added multiple layers to the structure of biological organization.

We have clearly not reached the bottom yet, and even once (if?) we do identify all the molecular players, a full functional mapping of their interactions, including the weak non-primary relationships, is certainly still well in the future. Moreover, network structure itself may not be a fixed property of an organism. Recent developments have shown that the structure of protein-protein interaction (PPI) networks correlates with cell type (Huttlin et al., 2020). In this study, the authors of the BioPlex network have constructed PPI networks of two different cell types using the method described above and observed largely the same proteins engaged in different interactions. In other words, rewiring of PPI networks correlated with system state, adding another major challenge and level of complexity in our ability to infer biological network structure.

In turn, this lack of ability to fully map biochemical networks could be an important barrier for our ability to apply key tools from complex systems theory—notably network theory/graph theory—to these networks. Of course, we do not mean to imply that network theory has no applications in biology. Indeed, there will be clear applications in the simplest systems

(Nijhout et al., 2017), or when emergent phenomena are clear enough to create higher-order networks that can be mapped (Bashan et al., 2012). But while the ability to predict and control nematode behavior through a network of 40 neurons is on the one hand an excellent and impressive demonstration of the power of network theory (Yan et al., 2017), it is also a humbling demonstration of how much more would be required to achieve similar prediction/control of human behavior, particularly when the number of neurons and their connection structure is not fixed from one individual to another.

In short, we argue that most models of biological networks are—and, for the foreseeable future, are likely to remain—*sparsely sampled systems*, systems in which the nodes and their edges are insufficiently known or sampled to robustly and directly characterize network-level properties. In such systems, insights will need to come not through exhaustive data, but through judicious use of methods to extract the relevant signals despite the sparseness of the data. This is true despite the advent of massive -omics, and despite the data processing tools that are being developed to manage that information.

DATA TOOLS FOR SPARSELY SAMPLED NETWORKS

While the above situation might sound pessimistic with regard to our potential to make progress in understanding biology, it is anything but. In fact, the same system properties that make biology so complex and hard to map also make it potentially more tractable as a complex system with the right analytical tools. It is the goal-directedness of biological systems that gives their components teleological functions: hearts are for pumping blood through the circulatory system, B-cells are for generating the antibodies in the adaptive immune response, and ribosomes are for translating mRNA into proteins (teleological is used here in the sense that purposeful behavior of systems can arise from purposeless evolution via natural selection, a shorthand that allows reduction of otherwise intractable, purely descriptive, mechanistic models). There is no equivalent set of functions in, say, an ecological network: what are flower-pollinator relationships, or plankton communities in marine ecosystems, "for"? And indeed, rightly or wrongly, for centuries biologists have relied on this kind of teleological understanding of biological components without a second thought.

More generally, we can conceive of most biological regulation as adjusting the condition of the organism to maintain broad-sense homeostasis. This regulation can occur across two types of dimension: discrete and continuous (**Figure 2**). For example, migratory birds arriving for the breeding season go through a discrete transition to a breeding state (Jacobs and Wingfield, 2000; Williams, 2012). This involves not just changes to sex hormones and gonad size, but also changes in immune function, diet, metabolism, and likely a host of other aspects of physiology that have not been studied extensively. We can thus conceive of breeding vs. non-breeding state as two "attractor states" of the system, in complex systems terminology. Small changes in physiology within a state are generally not sufficient to make the
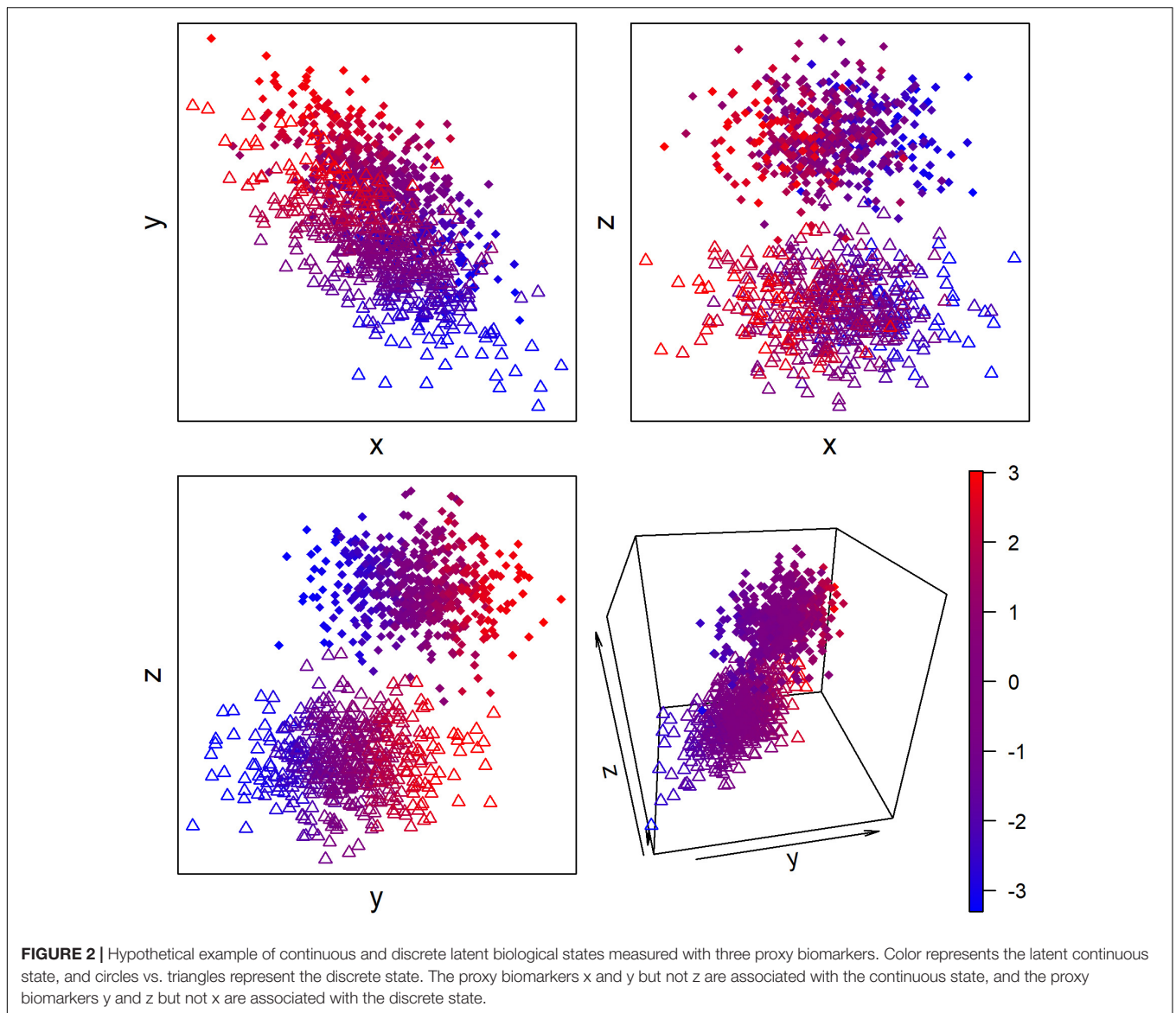
organism jump from one "attractor basin" to the other—only the right, and sufficient, stimulus can affect the change. Other discrete states could include sleeping vs. wakefulness, an immune system fighting acute infection, or the dauer stage of *C. elegans*.

Continuous variation occurs along a gradient. The chronic low-grade inflammatory response characteristic of mammalian aging is a good example (Franceschi et al., 2000), as is hunger, or the metabolic changes that accompany continuously increasing levels of exercise intensity. Both continuous and discrete processes are reflected in multiple, connected aspects of biology. Biological changes rarely happen in a vacuum, but involve coordination across multiple components that need to take into account what other parts of the system are doing (Csete and Doyle, 2004). And this is the property of biological systems that opens them up to tractability: the *coherence* of the overall system state. A biological system is, at any moment, in a state that reflects all the discrete and continuous dimensions of which it is composed: for example, a bird might be breeding, experiencing an acute infection of type X, under conditions of abundant food, while resting, etc. We do not yet have a complete enumeration of all these dimensions for any organism, and perhaps not even for any biological component. And these dimensions may not be additive: the presence of one state may preclude or modify others, and many of the biochemical components that help determine states are in fact shared. For example, the inflammatory molecule C-reactive protein (CRP) is involved in both acute infections, and in low-grade chronic inflammation associated with aging. CRP level is thus not a simple measure of either of these processes, but must be integrated with other information to give information about either (Bandeen-Roche et al., 2009; Morrisette-Thomas et al., 2014).

Broadly speaking, we can summarize the situation as follows: we have a set of biomarkers (taken here to mean any indicators of biological state, but likely individual molecules in biological networks), and we would like to make robust inferences about the state of the organism/cell/other biological component (let's call it an organism for simplicity). On the one hand, we have the threefold challenges of (1) imperfect knowledge of the states the organism might be in, (2) imperfect knowledge of how the biomarkers might interact with each other and/or overlap in how they signal different states; and (3) a marked undersampling of the relevant molecules (we assume). On the other hand, the associations among the biomarkers should reflect the structure of the underlying networks and the covariation in the biomarkers this generates: the coherence of the system state. Below we discuss several statistical approaches that can be used to extract information from such ensembles of biomarkers representing sparsely sampled networks.

Principal Components Analysis

Principal components analysis (PCA) is a data reduction technique that reorganizes the information in a large set of semi-redundant variables, permitting the user to extract a (much) smaller number of variables that explain the bulk of the variation in the original set. It is related to many other methods, such as factor analysis and t-SNE (Van Der Maaten and Hinton, 2008), that can obtain similar objectives through

FIGURE 2 | Hypothetical example of continuous and discrete latent biological states measured with three proxy biomarkers. Color represents the latent continuous state, and circles vs. triangles represent the discrete state. The proxy biomarkers x and y but not z are associated with the continuous state, and the proxy biomarkers y and z but not x are associated with the discrete state.

slightly different approaches. Our discussion here applies equally to all these methods.

While PCA is primarily considered or used as a way to reduce the number of dimensions in a dataset, it is also a powerful tool to understand those dimensions. For example, application of PCA to epidemiological data on malaria in India showed that seven indices could be effectively reduced to one, not two, PC axes (Cohen et al., 2010). This meant that indices showing the balance between the two main malaria species and those showing the abundance were actually moving largely in tandem, implying an ecological gradient dynamic in which one species dominates when abundance is high, and another dominates when abundance is low. Similarly, PCA analysis of inflammatory markers in human cohort data has shown that a single key axis describes much of the variation in a population of largely older adults, and that pro- and anti-inflammatory markers positively co-vary along this axis: chronic inflammation

is not characterized by high pro-inflammatory markers and low anti-inflammatory markers, but by high levels of both, an activation of the system in which the anti- markers chase the pro- markers without ever really catching up (Bandeen-Roche et al., 2009; Morrisette-Thomas et al., 2014; Varadhan et al., 2014; Cohen et al., 2018a). Similarly, PCA applied to standard clinical biomarkers has revealed a surprisingly stable structure that integrates multiple physiological systems (Cohen et al., 2015b). The first principal component is multi-systemic, with high scores indicating anemia, low calcium, low protein transport (e.g., albumin), and high chronic inflammation, and was termed "integrated albunemia." Integrated albunemia scores increase with age and predict mortality risk and clinical frailty but not chronic diseases. They are also elevated acutely during hospitalizations, etc. The correlation structure among the main implicated systems appears stable, such that integrated albunemia is a more robustly measurable phenomenon than its component

markers. For example, its correlation with age and disease is more stable than that of the component markers. This appears to be a good example of a continuously varying attractor state.

While PCA, when used appropriately, can be a powerful tool to uncover structure in biological data, it needs to be used with caution. In particular, it can be sensitive to population composition, and structure can also vary depending on conditions or population (Pigeon et al., 2013). For example, application of PCA to 13 circulating biomarkers in wild European starlings (*Sturnus vulgaris*) showed that the correlation structure—and thus the axes one might extract—varied across years and breeding season (Fowler et al., 2018). This is to be expected: changes in food abundance, diet composition, or parasite burden should be expected to vary across seasons and years, and also to cause changes in which biomarkers correlate with which. But if the correlation structure is unstable, how can we extract axes? One answer is that partial correlation structures can be extracted. For example, in bighorn sheep (*Ovis canadensis*), milk composition varies across both years and across individual mothers, and it was possible to extract a correlation structure for each of these aspects using appropriate hierarchical models (Renaud et al., 2019).

However, more generally, care should be used to apply PCA and related methods with appropriate cross-population validation to ensure that any findings are true reflections of biological organization rather than artifacts of sampling, population composition, or environmental heterogeneity (Cohen et al., 2018b). In the case of the Fowler et al. (2018) study, no meaningful axes could be extracted due to the strong changes in correlation structure and the limited samples sizes in the relevant subgroups to robustly estimate structure. In contrast, integrated albunemia has been validated in multiple populations (Cohen et al., 2015b) and even species (Wey et al., 2019). Likewise, there have multiple validations of inflamm-aging as a continuous process that can be identified by PCA and related methods (Bandeen-Roche et al., 2009; Cohen et al., 2015b, 2018a). The simplest test is to repeat the PCA in distinct populations or population subsets, extract the loadings, and then cross-apply them to the other populations to generate scores. This generates multiple versions of the PCA scores as calibrated based on, say, men, women, population 1, population 2, etc. A correlation matrix or correlogram can then be used to assess how well the same axis is extracted. In the case of integrated albunemia, for example, these correlations are generally greater than 0.95, even from populations on different continents. Further confirmation can be obtained by ensuring that the interpretation of the axis via its loadings is similar. Graphical approaches to this can be found in Cohen et al. (2015b).

Statistical Distance

Statistical distances are ways of quantifying how different an individual or group is from another individual or group, usually across a series of variables. In the context of measuring biological or physiological state, the principal application has been to quantify how different an individual's multivariate biomarker state is from some reference state, often the population average or some reference healthy state. This requires statistical distances that can quantify a distance between an individual and a group. The two main methods that do this are Euclidean distance and Mahalanobis distance (De Maesschalck et al., 2000). Euclidean distance does not take into account any correlations between the biomarkers, and thus, if redundant markers are included, will double-count that information. Mahalanobis distance assumes multivariate normality, and then uses the inverse of the correlation matrix to eliminate redundancies among highly correlated variables. This has the effect of down-weighting redundant variables.

Mahalanobis distance has, accordingly, been more widely used to measure physiological state. To our knowledge, the first application was to measure physiological declines prior to death in fruit flies (*Drosophila melanogaster*) (Shahrestani et al., 2012). It was subsequently and independently applied to human biomarker data for similar purposes, and has also been applied in wild animals, both with (Milot et al., 2014) and without (Fowler et al., 2018) success. The assumption in all these cases is that average state is close to optimal state, and that individuals far from optimal are likely to be more unhealthy. This assumption is related to theory suggesting that homeostatic states are relatively homogeneous, whereas there are numerous ways that homeostasis can be lost and thus a diversity of ways to diverge from the norm (Cohen, 2016). Accordingly, Mahalanobis distance has been proposed as a measure of homeostatic or physiological dysregulation. This proposition makes a number of predictions. First, Mahalanobis distance should increase with age. Second, it should predict a wide variety of adverse health outcomes after controlling for age. Third, this signal should not depend strongly on any single biomarker, implying that (a) the choice of biomarkers to include is not crucial, and (b) the signal should increase monotonically (but with diminishing marginal returns) as the number of biomarkers is increased. All of these predictions have now been confirmed by multiple studies (Cohen et al., 2013, 2014, 2015a, 2018c; Arbeev et al., 2016; Dansereau et al., 2019; Kraft et al., 2020; Liu, 2020). The third prediction is particularly important from a complex systems perspective, as it implies that the signal is diffuse or distributed in the physiological networks, and thus robustly estimable from subsamples of biomarkers even without detailed knowledge of network structure. It does thus appear that Mahalanobis distance is a valid metric of homeostatic dysregulation, though it certainly involves substantial measurement error. However, an advantage of the Mahalanobis distance approach is that it makes no prior assumptions about "good" or "bad" levels of biomarkers, and does not calibrate based on age or anything else (unless a specific reference population is chosen). This makes it agnostic and neutral for subsequent physiological inferences.

Mahalanobis distance can also be applied to particular physiological systems by dividing biomarkers into groups based on *a priori* knowledge. Indeed, *a priori* knowledge does a reasonably good job of distinguishing systems with minimally correlated dysregulation levels (Li et al., 2015). Net of age, correlations among physiological systems measurable with standard clinical human biomarkers are generally significant but weak ($r < 0.2$), implying feedback effects among the systems.

This led to the prediction that similar types of systems might be identified in -omics data, notably DNA methylation, gene expression, or proteomics. However, this has not yet been proven, and indeed in gene expression data from human blood samples, the opposite pattern emerged: most systems identified via gene ontology did not show any significant correlations with age, and those that did show correlations that replicated across datasets were uniformly negative rather than positive (Dufour et al., 2020). This might be indicative of these systems losing responsiveness to stimuli with age, but that remains speculation at this point.

Cluster Analysis

While PCA and Mahalanobis distance are both good measures of continuous processes, cluster analysis is more appropriate for detecting discrete states. Like PCA and Mahalanobis distance, many clustering algorithms are largely agnostic/uncalibrated/unsupervised, generating clusters based on similarities or differences in the data rather than an external target. There are many clustering algorithms, each with strengths and weaknesses, and a review is beyond the scope of this article (Scheibler and Schneider, 1985; Budayan et al., 2009). Hierarchical clustering, for example, has been profitably applied to biomarker data (Sebastiani et al., 2017). In some sense, clustering can be thought of as a discretized version of PCA, for when the target process states are discrete rather than continuous. A challenge here is that it is not always possible to know whether the states/processes of interest are continuous or discrete *a priori*. In fact, in some cases, the set of biomarkers in question may capture both discrete and continuous processes (**Figure 2**). When there is doubt, it may be advisable to try both methods, with the objective of assessing the discreteness of the phenomena in question. Because of measurement error, biologically discrete processes may in fact appear overlapping, so the criterion for use of cluster analysis should be the presence of clear, though potentially overlapping, aggregates in multidimensional space. Fuzzy clustering methods may be appropriate in such cases (Budayan et al., 2009). Also, given that biological networks are both weighted and directed, it may be relevant to apply clustering methods specifically designed for such networks (Clemente and Grassi, 2018; Barajas-Martínez et al., 2020).

A second challenge with cluster analysis is the diversity of methods available. These methods often give discordant results, identifying clusters that are not necessarily similar in their composition from one method to the next. This is, to some extent, to be expected. If we choose a group of 100 people and try to cluster them based on similarity, our groups could look very different if we base the similarity on demographics, on health state, on music preferences, etc. There is no "true" way to cluster the people. Even with the same set of variables, the way they are weighted and treated should be expected to have an impact. Nonetheless, if the clusters reflect true biologically discrete states, we should expect a certain reproducibility across methods. In this sense, cross-method validation of cluster analysis could prove an important tool for identifying important biological attractor states.

Statistical Network Inference

It is possible to combine analysis of correlation structure and clustering methods to infer network structure. This has been done both with general physiological biomarkers (Barajas-Martínez et al., 2020), and with -omics data, for example using the weighted gene coexpression network analysis (WGCNA, Zhang and Horvath, 2005; Hariharan et al., 2014). The idea here is that the correlation structure of variables reveals the network structure, and in the context of high-dimensional -omics data may also reveal discrete modules. We view this approach as one with great potential, but also one that needs to be used with caution. In particular, substantial work needs to be done to assess the robustness of the estimation of network structure to data stochasticity. The Cohen lab has unpublished analyses on gene expression data in which WGCNA failed to produce even minimally similar structures across different datasets; this may be due to the difficulty correcting for batch effects and how this impacts correlation structure, or to population differences in correlation structure (Dufour et al., 2020). It is well-known that correlation structures do change, across age (Barajas-Martínez et al., 2020), across environment (Fowler et al., 2018), and across cell type (Huttlin et al., 2020), among others. Such instability of the basic network structure could be problematic, but also could be the feature of interest (Fowler et al., 2018; Barajas-Martínez et al., 2020). It is also worth noting that such methods generally characterize the network structure of a population, but do not permit direct evaluation of an individual's physiological state. If network structure is itself malleable, network structure may simply be one more indicator of the attractor state of the individual, in which case methods that quantify individual state could serve as a proxy for network structure after appropriate validation. This could open up many potential research directions.

Machine Learning

PCA, Mahalanobis distance, and cluster analysis are relatively neutral methods that make few assumptions about the data. They do need to be used in biologically informed ways—for example, choice of variables and appropriate cross-population validation for PCA, and choice of reference population for Mahalanobis distance—but beyond this they are largely agnostic, or, in machine learning terms, unsupervised. In contrast, there are a host of supervised machine learning methods that have arisen in the last decades that permit the generation of precise algorithms to predict targets. For example, chronological age or measures of health status have been used to train random forests and deep neural nets applied to clinical biomarkers (Putin et al., 2016; Bello and Dumancas, 2017), and elastic nets or other regression-linked approaches applied to DNA methylation data and metabonomics data (Horvath, 2013; Hertel et al., 2016). However, such methods present substantial challenges when there is not a clear biological framework to link the biomarkers to the prediction target (Hertel et al., 2019), a symptom of a much more general challenge in artificial intelligence (Mitchell, 2020). These techniques are thus undoubtedly powerful, but work best when the biological nature of the prediction target is crystal clear,

and are thus of limited interest in the context of this review. That said, there are many applications of the newer machine learning methodologies that can be applied in an unsupervised way for the characterization of complex biological systems. They are too numerous to enumerate here, as our objective is less an exhaustive review than a demonstration of principle which we hope our readers will take in new directions.

DISCUSSION

We have argued here that a complex systems perspective on biological structure leads to reasons for both pessimism and optimism in terms of our ability to measure important biological states. On the pessimistic side, we are a long way from even identifying all the key molecules in the relevant biological networks, much less a full map of their pairwise interactions. Furthermore, the complexity of the underlying networks (and experience) both indicate that single biomarkers will often be generally poor indicators of system state (Cohen et al., 2018b). This is partly because the underlying networks are generally structured to adjust multiple outputs as a function of multiple inputs, requiring a regulatory network structure that involves multiple intermediary pathways that balance the potentially competing signals of the inputs, much as in a neural network or autoencoder (Csete and Doyle, 2004; Cohen et al., 2020). Because of this, it is rare that a single molecule has a universal interpretation. For example, interleukin-6 (IL-6) is widely considered the best marker of chronic inflammation. However, there are different types of inflammation, and IL-6 can even have anti-inflammatory roles in some contexts (Bandeen-Roche et al., 2009). It is a decent marker on average, but performs worse than multivariate approaches designed to integrate across numerous inflammatory cytokines (Cohen et al., 2018a).

On the optimistic side, the basic coherence of biological function implies that key biological states should be measurable with small ensembles of molecules, often arbitrarily chosen within the sphere of general interest. Full knowledge of the network is not needed to make progress. This assertion bears some qualitative similarities to the field of compressed sensing (Donoho, 2006), though we note that in that case there is an assumption of "sparseness" in a technical sense (i.e., most coefficients are zero), whereas we have referred to sparseness not of coefficients, but of data, and in a more colloquial sense. We thus do not believe, though cannot yet prove, that there is no formal link between our assertions here and compressed sensing.

Between the optimism and the pessimism lies the notion of biological function, the teleological glue that holds together most of our understanding of biology. Physical structures (e.g., tissue organization) and network structures (e.g., biochemical pathways) are thought to have functions, in some cases patently obvious, in some cases discovered through careful research, but in many cases still obscure. The approaches outlined here are examples of ways to elucidate network structures that may have gone unremarked, and in particular to link such

structures to functions. For example, if cluster analysis of single-cell gene expression reveals three stable yet heretofore unremarked profiles of what was previously thought to be a homogeneous cell type, there would be good reason to conduct further research on these three types in order to elucidate their respective roles.

In some sense, this assertion is completely unremarkable—of course we would explore the functions of such cell types if we found good evidence of distinctions, and cluster analysis is hardly new—and yet it also opens the door to a new framework for how we should explore biological variation. To date, we have tried to fit biological variation into our notions of how biology is organized, with varying degrees of success. For example, as noted above, many machine learning approaches have been applied to various kinds of data in an effort to quantify the aging process, without questioning the assumption that there is an aging process that could be measured. However, the existence of "aging" as a biologically (as opposed to culturally) meaningful concept is now being questioned (Cohen et al., 2020). Occasionally when efforts to map data onto existing concepts fails, a new paradigm emerges from the data, such as the alternative proteins discussed above. But the judicious use of methods such as those described here offers the potential for much more: a systematic and data-driven exploration of the structure, and thus function, of biological organization.

In short, we are proposing a broad effort to map biological variability based on several principles:

(1) **Analyses should be largely data-driven and conceptually agnostic.** We often do not understand the key structures, states, and processes, and these may or may not map well to the concepts and words we employ to describe them. However, general biological knowledge should be used to structure the analyses. Note that the annotations underlying the data are reflections of our biological conceptions and assumptions (Brunet et al., 2018), so there is no perfectly agnostic approach.

(2) **Replicable patterns in the data indicate key aspects of biological organization.**

(3) **Identifying structures and patterns will point to function** (Han, 2008). And function, once determined, will give meaning to the findings.

(4) **Collect information on context**. For example, information on an organism's sex, age, environment, etc. could prove essential to uncovering the functional relevance of different states.

Note that the approach we propose also avoids one of the primary criticisms possible for a teleological or functional (as opposed to mechanistic) understanding of biology: that our conceptions of function are biased by cultural or other subjective factors. In fact, a data-driven approach to identifying functional biological units could help actively identify cases where subjective factors have unduly influenced our understanding, while simultaneously permitting us to harness the explanatory power of natural selection as a force that shapes networks to achieve certain functional objectives.

The methods illustrated above—PCA, statistical distance, and cluster analysis—are examples, not recommendations: many others could be employed, and new ones will be developed, particularly as we move toward integration of multi-omics. The key point is that even standard, well-known techniques can reveal hidden structures in biological data, allowing the data to tell us a story that is often quite different from what we expected. For a long time, biological research has been largely hypothesis driven, an approach that works well with smaller data scales and when every variable is carefully chosen and laboriously measured. But in the world of big data and -omics, we are faced with the scale of our ignorance and the impossibility of generating enough well-founded hypotheses. The approach we propose here offers a way forward, a basic mapping of the landscape of biological function through networks, on which future hypotheses can be built.

AUTHOR CONTRIBUTIONS

AC conceived of and wrote the main manuscript. SL created **Figure 1** and wrote sections on alternative proteins in consultation with XR. All authors revised and edited the full manuscript.

FUNDING

REFERENCES

Albert, R., and Barabási, A. L. (2002). Statistical mechanics of complex networks. *Rev. Mod. Phys.* 74, 47–97. doi: 10.1103/RevModPhys.74.47

Arbeev, K. G., Cohen, A. A., Arbeeva, L. S., Milot, E., Stallard, E., Kulminski, A. M., et al. (2016). Optimal versus realized trajectories of physiological dysregulation in aging and their relation to sex-specific mortality risk. *Front. Public Heal.* 4:3. doi: 10.3389/fpubh.2016.00003

Bandeen-Roche, K., Walston, J. D., Huang, Y., Semba, R. D., and Ferrucci, L. (2009). Measuring systemic inflammatory regulation in older adults: evidence and utility. *Rejuvenation Res.* 12, 403–410. doi: 10.1089/rej.2009.0883

Barabási, A. L. (2016). *Network Science*. Cambridge: Cambridge University Press.

Barabási, A. L., Gulbahce, N., and Loscalzo, J. (2011). Network medicine: a network-based approach to human disease. *Nat. Rev. Genet.* 12, 56–68. doi: 10.1038/nrg2918

Barajas-Martínez, A., Easton, J. F., Rivera, A. L., Martínez-Tapia, R., de la Cruz, L., Robles-Cabrera, A., et al. (2020). Metabolic physiological networks: the impact of age. *Front. Physiol.* 11:587994. doi: 10.3389/fphys.2020.587994

Bartsch, R. P., Liu, K. K. L., Bashan, A., and Ivanov, P. C. (2015). Network physiology: how organ systems dynamically interact. *PLoS One* 10:e0142143. doi: 10.1371/journal.pone.0142143

Bartsch, R. P., Liu, K. K. L., Ma, Q. D. Y., and Ivanov, P. C. (2014). Three independent forms of cardio-respiratory coupling: transitions across sleep stages. *Comput. Cardiol.* 41, 781–784.

Bashan, A., Bartsch, R. P., Kantelhardt, J. W., Havlin, S., and Ivanov, P. C. (2012). Network physiology reveals relations between network topology and physiological function. *Nat. Commun.* 3:702. doi: 10.1038/ncomms1705

Bello, G. A., and Dumancas, G. (2017). A random forest-induced distance-based measure of physiologic al dysregulation. *Curr. Aging Sci.* 10, 270–281.

Brunet, M. A., Brunelle, M., Lucier, J.-F., Delcourt, V., Levesque, M., Grenier, F., et al. (2019). OpenProt: a more comprehensive guide to explore eukaryotic coding potential and proteomes. *Nucleic Acids Res.* 47, D403–D410. doi: 10.1093/nar/gky936

Brunet, M. A., Leblanc, S., and Roucou, X. (2020). Reconsidering proteomic diversity with functional investigation of small ORFs and alternative ORFs. *Exp. Cell Res.* 393:112057. doi: 10.1016/j.yexcr.2020.112057

Brunet, M. A., Levesque, S. A., Hunting, D. J., Cohen, A. A., and Roucou, X. (2018). Recognition of the polycistronic nature of human genes is critical to understanding the genotype-phenotype relationship. *Genome Res.* 28, 609–624. doi: 10.1101/gr.230938.117

Budayan, C., Dikmen, I., and Birgonul, M. T. (2009). Comparing the performance of traditional cluster analysis, self-organizing maps and fuzzy C-means method for strategic grouping. *Expert Syst. Appl.* 36, 11772–11781. doi: 10.1016/j.eswa.2009.04.022

Chen, J., Brunner, A. D., Cogan, J. Z., Nuñez, J. K., Fields, A. P., Adamson, B., et al. (2020). Pervasive functional translation of noncanonical human open reading frames. *Science* 367, 1140–1146. doi: 10.1126/science.aav5912

Chow, L. T., Broker, T. R., and Lewis, J. B. (1979). Complex splicing patterns of RNAs from the early regions of adenovirus-2. *J. Mol. Biol.* 134, 265–303. doi: 10.1016/0022-2836(79)90036-6

Clemente, G. P., and Grassi, R. (2018). Directed clustering in weighted networks: a new perspective. *Chaos Solitons Fractals* 107, 26–38. doi: 10.1016/j.chaos.2017.12.007

Close, P., Gillard, M., Ladang, A., Jiang, Z., Papuga, J., Hawkes, N., et al. (2012). DERP6 (ELP5) and C3ORF75 (ELP6) regulate tumorigenicity and migration of melanoma cells as subunits of elongator. *J. Biol. Chem.* 287, 32535–32545. doi: 10.1074/jbc.M112.402727

Cohen, A. A. (2016). Complex systems dynamics in aging: new evidence, continuing questions. *Biogerontology* 17, 205–220. doi: 10.1007/s10522-015-9584-x

Cohen, A. A., Bandeen-Roche, K., Morissette-Thomas, V., and Fulop, T. (2018a). "A robust characterization of inflamm-aging and other immune processes through multivariate analysis of cytokines from longitudinal studies," in *Handbook on Immunosenescence: Basic Understanding and Clinical Applications*, eds T. Fulop, C. Franceschi, K. Hirokawa, and G. Pawelec (Doderecht: Springer Science + Business Media BV).

Cohen, A. A., Dhingra, N., Jotkar, R. M., Rodriguez, P. S., Sharma, V. P., and Jha, P. (2010). The summary index of malaria surveillance (SIMS): a stable index of malaria within india. *Popul. Health Metr.* 8:1. doi: 10.1186/1478-7954-8-1

Cohen, A. A., Legault, V., Fuellen, G., Fülöp, T., Fried, L. P., and Ferrucci, L. (2018b). The risks of biomarker-based epidemiology: associations of circulating calcium levels with age, mortality, and frailty vary substantially across populations. *Exp. Gerontol.* 107, 11–17. doi: 10.1016/j.exger.2017.07.011

Cohen, A. A., Legault, V., and Fülöp, T. (2020). What if there's no such thing as "aging"? *Mech. Ageing Dev.* 192:111344. doi: 10.1016/j.mad.2020.111344

Cohen, A. A., Legault, V., Li, Q., Fried, L. P., and Ferrucci, L. (2018c). Men sustain higher dysregulation levels than women without becoming frail. *J. Gerontol. Ser. A Biol. Sci. Med. Sci.* 73, 175–184. doi: 10.1093/gerona/glx146

Cohen, A. A., Li, Q., Milot, E., Leroux, M., Faucher, S., Morissette-Thomas, V., et al. (2015a). Statistical distance as a measure of physiological dysregulation is largely robust to variation in its biomarker composition. *PLoS One* 10:e0122541. doi: 10.1371/journal.pone.0122541

Cohen, A. A., Martin, L. B., Wingfield, J. C., McWilliams, S. R., and Dunne, J. A. (2012). Physiological regulatory networks: ecological roles and evolutionary constraints. *Trends Ecol. Evol.* 27, 428–435. doi: 10.1016/j.tree.2012.04.008

Cohen, A. A., Milot, E., Li, Q., Bergeron, P., Poirier, R., Dusseault-Bélanger, F., et al. (2015b). Detection of a novel, integrative aging process suggests complex physiological integration. *PLoS One* 10:e0116489. doi: 10.1371/journal.pone.0116489

Cohen, A. A., Milot, E., Li, Q., Legault, V., Fried, L. P., and Ferrucci, L. (2014). Cross-population validation of statistical distance as a measure of physiological dysregulation during aging. *Exp. Gerontol.* 57, 203–210. doi: 10.1016/j.exger.2014.04.016

Cohen, A. A., Milot, E., Yong, J., Seplaki, C. L., Fülöp, T., Bandeen-Roche, K., et al. (2013). A novel statistical approach shows evidence for multi-system physiological dysregulation during aging. *Mech. Ageing Dev.* 134, 110–117. doi: 10.1016/j.mad.2013.01.004

Crick, F. (1970). Central dogma of molecular biology. *Nature* 227, 561–563. doi: 10.1038/227561a0

Csete, M., and Doyle, J. (2004). Bow ties, metabolism and disease. *Trends Biotechnol.* 22, 446–450. doi: 10.1016/j.tibtech.2004.07.007

Dansereau, G., Wey, T. W., Legault, V., Brunet, M. A., Kemnitz, J. W., Ferrucci, L., et al. (2019). Conservation of physiological dysregulation signatures of aging across primates. *Aging Cell* 18:e12925. doi: 10.1111/acel.12925

De Maesschalck, R., Jouan-Rimbaud, D., and Massart, D. L. (2000). The Mahalanobis distance. *Chemom. Intell. Lab. Syst.* 50, 1–18. doi: 10.1016/s0169-7439(99)00047-7

Donoho, D. L. (2006). Compressed sensing. *IEEE Trans. Inf. Theory* 52, 1289–1306. doi: 10.1109/TIT.2006.871582

Dufour, F., Jacques, P. -É, and Cohen, A. A. (2020). Physiological system dysregulation in gene expression correlates negatively with age. *bioRxiv* [Preprint]. doi: 10.1101/2020.09.09.289389

Fowler, M. A., Paquet, M., Legault, V., Cohen, A. A., and Williams, T. D. (2018). Physiological predictors of reproductive performance in the European Starling (*Sturnus vulgaris*). *Front. Zool.* 15:45. doi: 10.1186/s12983-018-0288-3

Franceschi, C., Bonafé, M., Valensin, S., Olivieri, F., De Luca, M., Ottaviani, E., et al. (2000). Inflamm-aging: an evolutionary perspective on immunosenescence. *Ann. N. Y. Acad. Sci.* 908, 244–254. doi: 10.1111/j.1749-6632.2000.tb06651.x

Han, J.-D. J. (2008). Understanding biological functions through molecular networks. *Cell Res.* 18, 224–237. doi: 10.1038/cr.2008.16

Hariharan, R., Hoffman, J. M., Thomas, A. S., Soltow, Q. A., Jones, D. P., and Promislow, D. E. L. (2014). Invariance and plasticity in the *Drosophila melanogaster* metabolomic network in response to temperature. *BMC Syst. Biol.* 8:139. doi: 10.1186/s12918-014-0139-6

Harris, M. A., Clark, J., Ireland, A., Lomax, J., Ashburner, M., Foulger, R., et al. (2004). The Gene Oncology (GO) database and informatics resource. *Nucleic Acids Res.* 32, D258–D261. doi: 10.1093/nar/gkh036

Hertel, J., Frenzel, S., König, J., Wittfeld, K., Fuellen, G., Holtfreter, B., et al. (2019). The informative error: a framework for the construction of individualized phenotypes. *Stat. Methods Med. Res.* 28, 1427–1438. doi: 10.1177/0962280218759138

Hertel, J., Friedrich, N., Wittfeld, K., Pietzner, M., Budde, K., Van Der Auwera, S., et al. (2016). Measuring biological age via metabonomics: the metabolic age score. *J. Proteome Res.* 15, 400–410. doi: 10.1021/acs.jproteome.5b00561

Holland, J. H. (1992). Complex adaptive systems. *Daedalus* 121, 17–30. doi: 10.2307/20025416

Horvath, S. (2013). DNA methylation age of human tissues and cell types. *Genome Biol.* 14, 1–20.

Huttlin, E. L., Bruckner, R. J., Navarrete-Perea, J., Cannon, J. R., Baltier, K., Gebreab, F., et al. (2020). Dual proteome-scale networks reveal cell-specific remodeling of the human interactome. *bioRxiv* [Preprint]. doi: 10.1101/2020.01.19.905109

Ingolia, N. T., Lareau, L. F., and Weissman, J. S. (2011). Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. *Cell* 147, 789–802. doi: 10.1016/j.cell.2011.10.002

Ivanov, P. C. H., Liu, K. K. L., and Bartsch, R. P. (2016). Focus on the emerging new fields of network physiology and network medicine. *New J. Phys.* 18:100201. doi: 10.1088/1367-2630/18/10/100201

Ives, A. R. (1995). Measuring resilience in stochastic systems. *Ecol. Monogr.* 65, 217–233. doi: 10.2307/2937138

Jacobs, J. D., and Wingfield, J. C. (2000). Endocrine control of life-cycle stages: a constraint on response to the environment? *Condor* 102, 35–51. doi: 10.2307/1370406

Kanehisa, M., and Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 28, 27–30. doi: 10.1093/nar/28.1.27

Kitano, H. (2002). Systems biology: a brief overview. *Science* 295, 1662–1664. doi: 10.1126/science.1069492

Kraft, T., Stiegilitz, J., Trumble, B. C., Garcia, A., Kaplan, H. S., and Gurven, M. D. (2020). Physiological dysregulation and aging in evolutionary perspective. *Philos. Trans. R. Soc. B Biol. Sci.* 375:20190610.

Krämer, A., Green, J., Pollard, J., and Tugendreich, S. (2014). Causal analysis approaches in ingenuity pathway analysis. *Bioinformatics* 30, 523–530. doi: 10.1093/bioinformatics/btt703

Kriete, A. (2013). Robustness and aging—a systems-level perspective. *Biosystems* 112, 37–48. doi: 10.1016/j.biosystems.2013.03.014

Leblanc, S., and Brunet, M. A. (2020). Modelling of pathogen-host systems using deeper ORF annotations and transcriptomics to inform proteomics analyses. *Comput. Struct. Biotechnol. J.* 18, 2836–2850. doi: 10.1016/j.csbj.2020.10.010

Leblanc, S., Brunet, M. A., Lekehal, A. M., Duclos, A., Tremblay, A., Bruggeman-Gascon, A., et al. (2020). New-found coding potential of transcripts unveils missing members of human protein communities. *bioRxiv* [Preprint]. doi: 10.1101/2020.12.02.406710

Li, Q., Wang, S., Milot, E., Bergeron, P., Ferrucci, L., Fried, L. P., et al. (2015). Homeostatic dysregulation proceeds in parallel in multiple physiological systems. *Aging Cell* 14, 1103–1112. doi: 10.1111/acel.12402

Liu, Z. (2020). Development and validation of two composite aging measures using routine clinical biomarkers in the Chinese population: analyses from two prospective cohort studies. *J. Gerontol. Ser. A* glaa238. doi: 10.1093/gerona/glaa238

Milot, E., Cohen, A. A., Vézina, F., Buehler, D. M., Matson, K. D., and Piersma, T. (2014). A novel integrative method for measuring body condition in ecological studies based on physiological dysregulation. *Methods Ecol. Evol.* 5, 146–155. doi: 10.1111/2041-210X.12145

Mitchell, M. (2020). On crashing the barrier of meaning in artificial intelligence. *AI Mag.* 41, 86–92. doi: 10.1609/aimag.v41i2.5259

Mobus, G. E., and Kalton, M. C. (2014). *Principles of Systems Science*. New York, NY: Springer.

Morrisette-Thomas, V., Cohen, A. A., Fülöp, T., Riesco, T., Legault, V., Li, Q., et al. (2014). Inflamm-aging does not simply reflect increases in pro-inflammatory markers. *Mech. Ageing Dev.* 139, 49–57. doi: 10.1016/j.mad.2014.06.005

Navarro-Hernandez, I. C., López-Ortega, O., Acevedo-Ochoa, E., Cervantes-Díaz, R., Romero-Ramírez, S., Sosa-Hernández, V. A., et al. (2020). Tetraspanin 33 (TSPAN33) regulates endocytosis and migration of human B lymphocytes by affecting the tension of the plasma membrane. *FEBS J.* 287, 3449–3471. doi: 10.1111/febs.15216

Nemeth, E., Valore, E. V., Territo, M., Schiller, G., Lichtenstein, A., and Ganz, T. (2003). Hepcidin, a putative mediator of anemia of inflammation, is a type II acute-phase protein. *Blood* 101, 2461–2463. doi: 10.1182/blood-2002-10-3235

Nijhout, H. F., Sadre-Marandi, F., Best, J., and Reed, M. C. (2017). Systems biology of phenotypic robustness and plasticity. *Integr. Comp. Biol.* 57, 171–184. doi: 10.1093/icb/icx076

Pigeon, G., Bélisle, M., Garant, D., Cohen, A. A., and Pelletier, F. (2013). Ecological immunology in a fluctuating environment: an integrative analysis of tree swallow nestling immune defense. *Ecol. Evol.* 3, 1091–1103. doi: 10.1002/ece3.504

Podobnik, B., Jusup, M., Tiganj, Z., Wang, W. X., Buldú, J. M., and Stanley, H. E. (2017). Biological conservation law as an emerging functionality in dynamical neuronal networks. *Proc. Natl. Acad. Sci. U.S.A.* 114, 11826–11831. doi: 10.1073/pnas.1705704114

Podobnik, B., Korošak, D., Skelin Klemen, M., Stožer, A., Dolenšek, J., Slak Rupnik, M., et al. (2020). β cells operate collectively to help maintain glucose homeostasis. *Biophys. J.* 18, 2588–2595. doi: 10.1016/j.bpj.2020.04.005

Putin, E., Mamoshina, P., Aliper, A., Korzinkin, M., Moskalev, A., Kolosov, A., et al. (2016). Deep biomarkers of human aging: application of deep neural networks to biomarker development. *Aging (Albany NY)* 8, 1021–1033. doi: 10.18632/aging.100968

Renaud, L. A., Blanchet, F. G., Cohen, A. A., and Pelletier, F. (2019). Causes and short-term consequences of variation in milk composition in wild sheep. *J. Anim. Ecol.* 88, 857–869. doi: 10.1111/1365-2656.12977

Samandi, S., Roy, A. V., Delcourt, V., Lucier, J.-F., Gagnon, J., Beaudoin, M. C., et al. (2017). Deep transcriptome annotation enables the discovery and functional characterization of cryptic small proteins. *Elife* 6:e27860. doi: 10.7554/eLife.27860

Scheibler, D., and Schneider, W. (1985). Monte carlo tests of the accuracy of cluster analysis algorithms: a comparison of hierarchical and nonhierarchical methods. *Multivariate Behav. Res.* 20, 283–304. doi: 10.1207/s15327906mbr2003_4

Scott, M. S., and Ono, M. (2011). From snoRNA to miRNA: dual function regulatory non-coding RNAs. *Biochimie* 93, 1987–1992. doi: 10.1016/j.biochi.2011.05.026

Sebastiani, P., Thyagarajan, B., Sun, F., Schupf, N., Newman, A. B., Montano, M., et al. (2017). Biomarker signatures of aging. *Aging Cell* 16, 329–338. doi: 10.1111/acel.12557

Shahrestani, P., Tran, X., and Mueller, L. D. (2012). Physiology declines prior to death in *Drosophila melanogaster*. *Biogerontology* 13, 537–545. doi: 10.1007/s10522-012-9398-z

Spiegelman, S., Watson, K. F., and Kacian, D. L. (1971). Synthesis of DNA complements of natural RNAs: a general approach. *Proc. Natl. Acad. Sci. U.S.A.* 68, 2843–2845. doi: 10.1073/pnas.68.11.2843

Sterling, P. (2015). "Principles of allostasis: optimal design, predictive regulation, pathophysiology, and rational therapeutics," in *Allostasis, Homeostasis, and the Costs of Physiological Adaptation*, ed. J. Schulkin (Cambridge: Cambridge University Press), doi: 10.1017/CBO9781316257081.004

Strogatz, S. H. (2001). Exploring complex networks. *Nature* 410, 268–276. doi: 10.1038/35065725

Tieri, P., Grignolio, A., Zaikin, A., Mishto, M., Remondini, D., Castellani, G. C., et al. (2010). Network, degeneracy and bow tie integrating paradigms and architectures to grasp the complexity of the immune system. *Theor. Biol. Med. Model.* 7:32. doi: 10.1186/1742-4682-7-32

Ukraintseva, S., Yashin, A. I., and Arbeev, K. G. (2016). Resilience versus robustness in aging. *J. Gerontol. Ser. A Biol. Sci. Med. Sci.* 71, 1533–1534. doi: 10.1093/gerona/glw083

Van Der Maaten, L., and Hinton, G. (2008). Visualizing data using t-SNE. *J. Mach. Learn. Res.* 9, 2579–2605.

Vanderperre, B., Lucier, J.-F., Bissonnette, C., Motard, J., Tremblay, G., Vanderperre, S., et al. (2013). Direct detection of alternative open reading frames translation products in human significantly expands the proteome. *PLoS One* 8:e70698. doi: 10.1371/journal.pone.0070698

Varadhan, R., Yao, W., Matteini, A., Beamer, B. A., Xue, Q., Yang, H., et al. (2014). Simple biologically informed inflammatory index of two serum cytokines predicts 10 year all-cause mortality in older adults. *J. Gerontol. Ser. A Biol. Sci. Med. Sci.* 69, 165–173. doi: 10.1093/gerona/glt023

Vinel, C., Lukjanenko, L., Batut, A., Deleruyelle, S., Pradère, J. P., Le Gonidec, S., et al. (2018). The exerkine apelin reverses age-associated sarcopenia. *Nat. Med.* 24, 1360–1371. doi: 10.1038/s41591-018-0131-6

Wagner, N. B., Weide, B., Gries, M., Reith, M., Tarnanidis, K., Schuermans, V., et al. (2019). Tumor microenvironment-derived S100A8/A9 is a novel prognostic biomarker for advanced melanoma patients and during immunotherapy with anti-PD-1 antibodies. *J. Immunother. Cancer* 7:343. doi: 10.1186/s40425-019-0828-1

Wey, T. W., Roberge, É., Legault, V., Kemnitz, J. W., Ferrucci, L., and Cohen, A. A. (2019). An emergent integrated aging process conserved across primates. *J. Gerontol. Sci. Biol. Sci. Med. Sci.* 74, 1689–1698. doi: 10.1093/gerona/glz110

Williams, T. D. (2012). *Physiological Adaptations for Breeding in Birds*. Princeton, NJ: Princeton University Press.

Yan, G., Vértes, P. E., Towlson, E. K., Chew, Y. L., Walker, D. S., Schafer, W. R., et al. (2017). Network control principles predict neuron function in the *Caenorhabditis elegans* connectome. *Nature* 550, 519–523. doi: 10.1038/nature24056

Zhang, B., and Horvath, S. (2005). A general framework for weighted gene co-expression network analysis. *Stat. Appl. Genet. Mol. Biol.* 4:17.