Chapter 14
Proteomic Network Systems Analysis

D. Kent Arrell and Andre Terzic

Abstract Proteomics and other high throughput technologies generate extensive molecular lists, the scope of which renders their accurate interpretation a daunting task. Thus, generalizable approaches by which to extract insight from this complexity are indispensable. Network systems biology principles and their application offer a modular, interchangeable data analytics pipeline by which to collate, integrate, and prioritize such datasets. By understanding the basis and utility of various organizing and interpretive profiling elements including ontological classification, functional enrichment and over-representation algorithms, and combining these with pathway analysis resources and the versatile tools and applications of complex network analysis, an applied network systems approach yields actionable insights into tackling the biology underlying high throughput data. Providing a framework to proteomic newcomers and experienced practitioners alike, we here outline data analytic approaches and provide concrete examples of the pairing of network systems prognostication with informed follow-up, through application of complementary physiological experimentation to validate proteomic observations in cardiovascular health and disease.

Keywords Bioinformatics • Cardiac • Cardiovascular • Complex network analysis • Heart disease • Network biology • Protein • Proteome • Systems biology

Introduction

The proteomics field incorporates a diverse array of methods and approaches by which to examine consequences of posttranscriptional and posttranslational effects on protein abundance, structure, modifications, and interactions. Many such elements are detailed in this manual, with chapters addressing various applications of

D.K. Arrell, PhD • A. Terzic, MD, PhD
Center for Regenerative Medicine and Marriott Heart Disease Research Program,
Division of Cardiovascular Diseases, Department of Medicine,
Department of Molecular Pharmacology and Experimental Therapeutics,
and Department of Medical Genetics, Mayo Clinic,
Rochester, MN, USA
e-mail: arrell.kent@mayo.edu; terzic.andre@mayo.edu

proteomic methodology, experimental design, data acquisition, and analysis of sub-cellular compartments and specific protein modifications to further our understanding of cardiovascular health and disease [1]. Regardless of this plethora of experimental and technical approaches, all proteomic studies share a common denominator: their output. Every study yields a compilation of proteins, either as simple lists of identities or subsets partitioned by differential expression or modification in response to the biological question under investigation. The key to extracting further insight from this output is the implementation of a generalized approach to data organization and interpretation.

From a logistical standpoint, this commonality – a list of proteins – suggests that it should be possible for any proteomic dataset to be examined in a similar manner. Unlike reductionist molecular approaches, where detailed functional analysis is conducted on a protein-by-protein basis, a more expansive strategy is required to account for the sheer volume of data. Indeed, while proteomic methods may be used to examine individual or small numbers of proteins, they are designed for and applied primarily to large scale analysis of entire proteomes or isolated subproteomes (Table 14.1), such that they now typically yield lists of several hundred to upwards of a thousand proteins in a single report. As one might anticipate, extracting insight from hundreds of proteins simultaneously is not facile, and can be perceived as altogether unmanageable. These lists can be simplified by narrowing the focus to a few choice proteins, such as those exhibiting the greatest extent of change, residing within a particular organelle, or executing a particular biological function. This offers advantages of reduced complexity and the potential to pinpoint functions germane to the topic of interest, but subjective protein exclusion during downstream analysis invariably leads to information loss. Moreover, care must be taken to ensure that data reduction decisions are not influenced by selection bias, whether intentional (e.g. confirmation bias [2], expectation bias [3]) or not. Finally, acquiring a mountain of data and ignoring all but the peak is patently counterproductive to an experimental rationale predicated on the capacity to conduct high throughput profiling. As we have emphasized previously [4], information reduction strategies may result in overlooking critical functional interactions and mechanistically important processes associated with and evident only upon examination of the complete dataset. In making these assessments, awareness of proteins not changing in a system is often just as valuable as detecting those that are altered. Thus, a judicious proteomic data analytics strategy should yield the same advantages – reduced information complexity and provision of functional insights – while remaining free of selection bias and ensuring inclusivity of all detected proteins [4, 5].

Network systems biology (Table 14.1) principles offer an attractive approach for comprehensive data analysis. Systems based strategies facilitate objective organization, prioritization, and integration of proteomic and other big data in their entirety, regardless of abundance, scope or complexity. To this end, a suite of bioinformatic and computational applications designed specifically to interrogate high throughput data is currently available, and in conjunction with protein databases that provide ease of access for molecular information retrieval, they harbor the capacity to bring clarity and cohesion to proteomic output [4–6]. Various applications and tools can

Table 14.1 Proteomic network systems analysis glossary

Subproteome	Subset of the proteome, often grouped by a shared cellular function, alteration by or dependency on a specific biological process, or components of a particular extracellular or intracellular organelle or compartment
Systems biology	Analytical high throughput data-inclusive approach to investigating and modeling relationships among a system's components in order to understand and predict emergent properties
Network or Interactome	Representation of pairwise collections of interactions and relationships, known as a graph in mathematics, which exhibits emergent properties based on their topological structure
Network systems biology	Complex network theory applied to the analysis of molecular interaction networks, including evaluation and assessment of their composition and topology to guide systems biological interpretation
Node	Individual subunits that comprise a graph or network via their cumulative interactions. In network biology, one or more of proteins, genes, metabolites, and endogenous small molecules or drugs are typically represented as nodes
Edge	An interaction or relationship between two nodes of a network. In network biology this might represent a physical complex of protein subunits, or a regulatory, genetic, or signaling effect, visualized as a line connecting interacting nodes. Based on the relationship, edges may also be portrayed as undirected or directed
Degree distribution	Node degree, the number of connections each node possesses to other nodes in a network, can be used to convey a network's degree distribution, which represents the probability distribution of the degrees of all network nodes
Scale-free topology	Non-stochastic power law distribution of node connectivity characteristic of biological networks, the non-random structure of which influences subsequent emergent network properties such as functional robustness
Hub	Network node that is much more highly connected (i.e. a greater degree) than would be expected to occur at random
Clustering coefficient	Extent to which nodes within a network cluster together as a measure of neighborhood relatedness, where a node's clustering coefficient is defined as the proportion of nodes connecting to it that also connect to each other
Bridging node	Node bridging the shortest direct path between two other nodes or modules within a network, often forming the shortest path between a high proportion of node pairs, leading to an increased likelihood of functional importance due to its position as a network bottleneck
Network motif	Patterns of complex network node-edge connections that are observed more frequently than expected relative to randomly assembled networks
Network module	Highly interconnected collection of nodes that nevertheless exhibits few interactions with the remaining network, often comprised of nodes with a shared function such as subunits of a protein complex or members of a metabolic or signaling pathway

Table 14.1 (continued)

Functional robustness	Network property of functional resilience when subjected to random removal or inhibition of a single node, based on the low connectivity of most nodes and the high degree of a limited proportion of nodes
Path length	Distance between two nodes in a network based on the minimum number of edges required to connect them. Average path length of scale-free networks is extremely small, meaning very few steps are required to connect any two nodes
Network diameter	Greatest path length required to connect any two nodes of a network
Metaboproteome	Subproteome comprised of and involving proteins supporting cell metabolism

be arranged and organized to suit particular projects and data sources, but in general there are four elements that form the basis of a proteomic network systems approach. The first component is ontological classification, an initial compartmentalization for partitioning high-throughput proteomic data into discrete biological categories. This serves to reduce complexity of an initial list of proteins, and enables assessment of the relative frequency or infrequency of occurrence for particular functional categories within acquired data relative to a specified reference set or between experimental cohorts. The second component, pathway analysis, extends categorical assignment by evaluating proteomic findings in the context of biological pathways. In this regard, data is superimposed onto canonical pathways and functional annotations, providing further evidence of enrichment properties and establishing connections between distinct elements of the measured dataset. Such connections, retrieved from pathway analysis resources or via stand-alone molecular interaction applications, are vital for the third component, complex network analysis. Networks, comprised of proteins connected by their collective functional and structural interactions, position proteins within the context of their local interaction neighborhood. Network composition, topology, and positional relevance of specific proteins provide value-added properties for data interpretation extending beyond what can be achieved with pathway analysis alone, and these can be exploited for hypothesis generation to assist in developing validation experiments to explain mechanistic underpinnings of initial proteomic measurements. A fourth and final optional component is systems modeling. If sufficient functional data is obtained relating to specific components, pathways, or network elements, it may be possible to develop mathematical or computational models to explain or predict functional outcomes on the basis of experimental proteomic findings. This particular element requires both extensive additional information to model the system under investigation as well as expertise in calculus and mathematics to design and implement. To simplify this introduction of proteomic network systems analysis for beginners, modeling will thus not be discussed here. The first three components, meanwhile, can be undertaken with only an initial list of output proteins and their related expression values, without any requisite expertise in standard tools of the trade. For those interested in further pursuing systems modeling, a recent description of fundamental concepts in its comprehension and application for cardiovascular proteomics is available [6].

Bioinformatic and computational network systems analytic approaches are prognostic in nature, providing hypothesis generating predictions requiring subsequent validation of underlying biological effects for the observed proteomic output. To avoid pitfalls of overestimating the impact of predictions or of statistical overfitting of high throughput data, predictive analytics should therefore be paired with complementary experimental data to ensure hypothesis validation. Ideally, then, prognosis should be actionable, and if predictions are valid, verifiable. This chapter provides a structured description of network systems biology procedures by which to address the daunting task of organizing and interpreting proteomic output, for generation of "actionable prognostication", and provides discrete examples of how information arising from these tools is paired with complementary physiological data to validate proteomic observations in cardiovascular research. The conceptual approach to proteomic network systems biology described herein extends just as readily to the analysis of genomic, transcriptomic, and metabolomic data. Thus, these organization and prioritization principles prove equally versatile for interrogation and interpretation of data from other high throughput molecular profiling methods, including strategies that integrate multi-level -omics datasets.

Ontological Classification, Functional Enrichment and Over-Representation

Current large-scale proteomic studies generate datasets exceedingly difficult to comprehend or interpret without initial data reduction or clustering. This is due to a combination of sheer size, i.e. the number of proteins, and an even greater complexity imposed by their associated biological functions and processes. In this regard, shared biological properties can and do serve as a useful starting point for data comprehension. As noted, when this process is conducted selectively to pare down an extensive list and focus immediately on a protein subset rather than the entire dataset, inclusivity is bypassed for simplicity and subsequent analysis is compromised by user bias. A more objective rationale involves collation of all detected proteins using extant biological information, with prioritization based on subsequent assessment and interrogation of the complete dataset. Even selection of statistical cutoffs should not be arbitrary, but rather applied as a reasonable fold change and statistical test with sufficient power to clearly establish a difference between experimental cohorts [7]. Implementing this strategy is aided by knowledge of the Gene Ontology (GO) [8], organized and structured specifically to document gene and protein properties, and awareness of where to find and access databases containing GO information.

In the field of proteomics, the UniProt Knowledge Base [9] (UniProtKB, accessible at www.uniprot.org) is an established protein sequence repository for assignment of spectral data acquired during mass spectrometry. Its current iteration emerged from a consortium combining several previous protein databases, including

Swiss-Prot, the Translated European Molecular Biology Laboratory, the Protein Information Resource and, more recently, the International Protein Index. Besides serving as a mass spectrometry resource, UniProtKB incorporates a wealth of additional information on various protein properties and characteristics, making it a useful starting point for in-depth dataset analysis [9]. Each entry cross-references with hyperlinks to a wealth of protein database resources listed under as many as 15 sub-categories, depending upon extent of protein characterization, and includes all available GO information [10]. As a result, initial ontology classification can be achieved by simply parsing GO data from UniProtKB, enabling data reduction to cluster proteins without necessitating in-depth or formal knowledge of the GO resource. In this way, proteins may be classified and grouped by a specific function (e.g. kinase, oxidoreductase), by particular biological processes to which they contribute (e.g. glycolysis, muscle contraction), where each protein might execute a unique function while working together as a collective within a particular pathway or as members of a multi-protein complex, and by their localization within one or more discrete cellular components (e.g. mitochondrion, nucleus).

It is precisely these concepts – molecular function, biological process, and cellular component – upon which the GO structure is based [11–16]. These are considered root terms forming 'domains' within the controlled vocabulary set out by the GO Consortium [8], and thus all other GO terms fall into one of these three domains. Each GO classification is unique, but together they form a loosely hierarchical structure, whereby more specialized 'child' terms link to one or multiple more generalized 'parent' terms. As such, the structure of related GO terms can be portrayed or described as a graph, or network, where every GO term serves as a node, interrelationships between pairs of GO terms form edges connecting them (Table 14.1), and edges in turn form nested connections within the hierarchy, the various elements of which fan out from their respective domain root terms (Fig. 14.1).

Comprehension of this structure is advantageous for individuals involved in high throughput research, as GO categorization is now applied to most cardiovascular proteomic studies [4, 6]. Clustering by shared functional properties would be straightforward if all proteins were defined by single GO terms. However, well characterized proteins are often assigned multiple GO associations, either as nested molecular functions of increasing specificity, as a result of participation in multiple biological processes, or a combination of the two, whereas a small proportion of proteins lack any GO term due to unknown function. Thus, a single protein may be included in and defined by multiple GO categories simultaneously (Fig. 14.1), or appear in none at all.

Following dataset assignment of protein GO designations, it is then beneficial to determine their categorical frequency of occurrence. On its own, the frequency at which a particular GO term appears in a dataset is somewhat meaningless. It must be interpreted in the context of an established benchmark such as the known proteome for a species or tissue of interest, or the full extent of proteins detected within the constraints of the experimental technology being applied, such as the complete set of proteins present on a chip array [12–19]. Statistical assessment is carried out by a hypergeometric distribution (Fig. 14.1), defining the probability of whether GO terms appear more or less frequently in an experimental dataset than would be
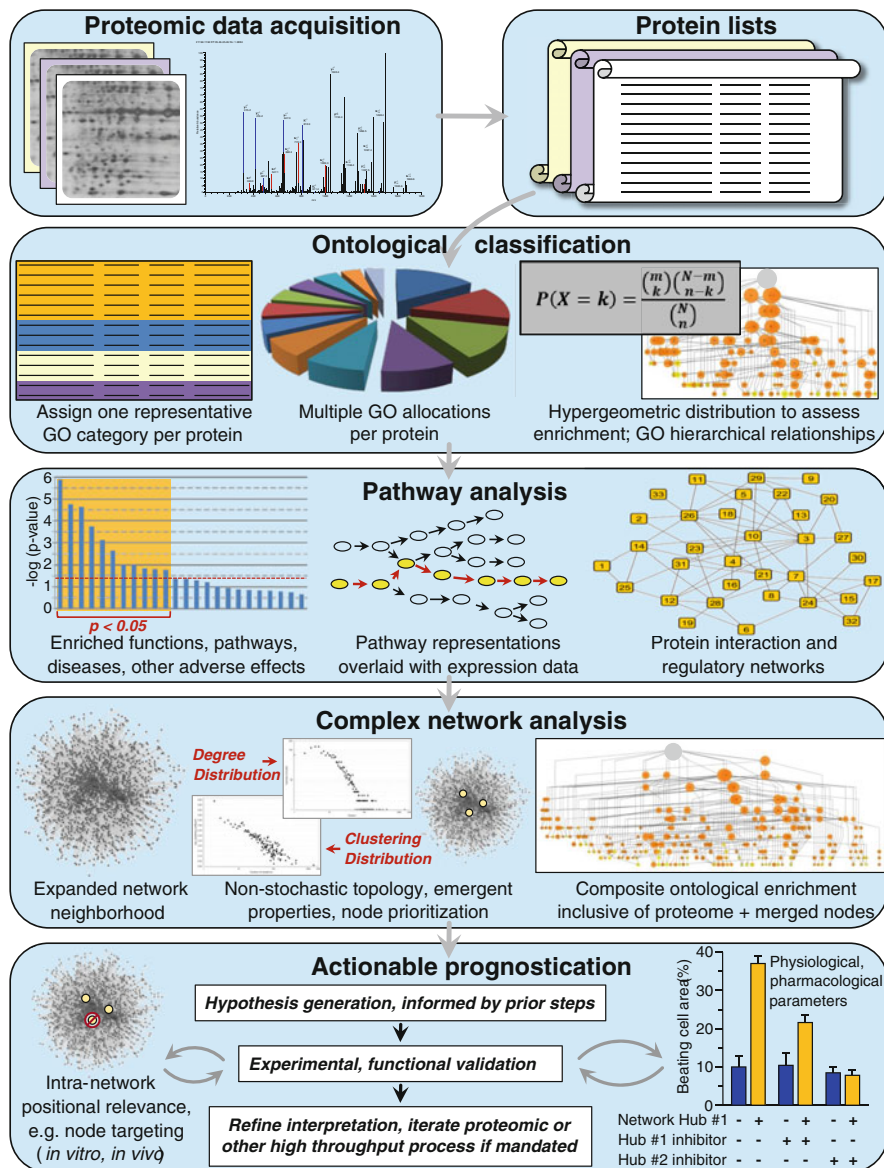
Fig. 14.1 Proteomic network systems analysis. Systems comprehension requires moving beyond simple lists by undertaking their organization, clustering and prioritization to reduce information complexity and extract functional insights into the biology underlying acquired proteomic data. Subjecting protein lists to ontological classification, pathway analysis, and complex network analysis, interpretation can be used to generate actionable hypotheses for functional validation. Shown as a simplified workflow, these elements do not necessarily follow a prescribed sequence, but instead serve as interchangeable components to be included as needed, or repeated multiple times at different stages of an analysis when warranted. Indeed, the entire process may be cyclical or iterative, as experimental validation and insights gleaned from an initial analysis may be further refined by subsequent proteomic or other high throughput data acquisition and a successive round of systems analysis. Abbreviation: *GO* Gene Ontology

anticipated relative to their occurrence within the reference benchmark set. It should be noted that overrepresentation and enrichment metrics are useful not only for differential expression analyses in terms of defining what classes of upregulated or downregulated proteins are overrepresented, but they are also applicable for examination of a simple list of protein identities. Here, enrichment analysis can be used to determine whether an experimental methodology is selective as desired or intended, or perhaps biased towards or against detection of a particular subproteome.

An alternative approach known as gene (or protein) set enrichment analysis can be used for differential comparison to previously published datasets [20–22]. This method takes a slightly different tack, avoiding prescribed statistical cutoffs for comparison, and instead makes use of rank ordered expression, which is then assessed for correlation or anti-correlation to the rank ordered expression of published data. This approach can be useful for teasing out subtle ontological differences between experimental groups, even in the absence of substantial numbers of significantly differing proteins. Fortunately, expertise in mathematics is not a prerequisite for conducting any of these tests as they are typically incorporated into various commercial and open source applications, including those used for pathway and network analysis.

Pathway Analysis

Efforts to extend proteomic categorization beyond the relative occurrence of, and assignment to, particular GO classifications are the domain of pathway analysis applications. Pathway algorithms facilitate expanded examination of proteomic data in the context of established biological pathways, protein functions, and their associated structural, functional and regulatory interactions. Proteomic datasets may be mapped across one or more specific canonical pathways, aiding in determination of whether particular pathways or pathway segments or branches are similarly or differentially affected, with mapping of proteins and expression level data onto pathways facilitating their visualization and representation. Many pathway applications enable protein annotation, using embedded information with details similar to those available within the UniProtKB, and tie-ins or hyperlinks to current knowledge on proteins of interest. As a consequence, many such algorithms also support enrichment and overrepresentation analysis, some making use of existing GO molecular function, biological process, and cellular component nomenclature, while other applications construct and implement platform-specific ontology terms and classifiers [4]. Finally, some pathway resources focus entirely on protein interactions, while others include interaction network generation as one element of a suite of functions, such as that offered by commercial pathway tools, e.g. MetaCore and Ingenuity Pathways Analysis (IPA). Network generation in these applications can then be further scrutinized for functional annotation enrichment or canonical pathway overrepresentation in the context of a broader network neighborhood in a fashion similar to that which is carried out on initial proteomic input. Beyond an

understanding of what pathway analysis entails, other primary issues that beginners are faced with are where to find these repositories, potential costs involved in their use, if any, and specified data formats, if required, all with the overarching consideration of what beneficial attributes are present and desirable in specific pathway analysis algorithms.

Even at first glance, it is evident that the extent of information available in individual pathway analysis resources varies greatly between applications. In part, this may be due to the fact that many pathway databases arose from investigator-generated data accumulation. Thus these databases may relate to a specific research area of interest, sometimes with a focus on only a limited set of pathways or a discrete protein property – such as protein-protein interactions – or an emphasis on data from only a single species or particular organelle. Other contributing factors are the sheer number of repositories, and their broad applicability to tackle the variability of biological questions under examination. The latest update of Pathguide, the largest online compendium of biological pathway and molecular interaction related resources, now lists nearly 550 different pathway applications [23]. This number has almost doubled in the past 3 years [4], signifying the tremendous growth in this field. Pathguide lists eight distinct pathway application categories: protein-protein interactions; metabolic pathways; signaling pathways; pathway diagrams; transcription factor/gene regulatory networks; protein-compound interactions; genetic interaction networks; protein sequence focused databases; and 17 separate resources listed under the category of 'Other'. This site provides detailed information about each resource, accessibility (cost and current availability), and whether they adhere to or accommodate specific bioinformatics language standards, e.g. BioPAX [24], which was designed to enable integration, exchange, visualization and analysis of biological pathway data. Many of these databases are freely accessible or free for use by academics, although the more comprehensive resources are typically commercial entities with license requirements that can be cost prohibitive to some investigators. In general, these resources map current biological knowledge to known pathways rather than serving as inference tools to predict theoretical interactions or novel biological outcomes. Some applications are attractive for use because of their broad applicability. For example, the MetaCore pathway database and the IPA Ingenuity Pathways Knowledge Base harbor suites of functions, accounting for these resources being listed in 6 and 5 Pathguide categories, respectively, making them popular choices for cardiovascular proteomic pathway analysis [11–14, 25–42]. Although less versatile, more specialized applications often prove highly desirable when matched to user-specific needs, for instance, exploiting mitochondrial protein interaction databases for bioenergetics research. For those with little knowledge of this bioinformatic field, Pathguide is an excellent source of information for an informed decision on selecting pathway analysis applications that optimally align with experimental needs.

Once a pathway analysis platform is selected, implementation may take many forms. A quick overview of IPA search parameters provides useful considerations in this regard. Input data can range from a number of high throughput experimental procedures, including proteomic, mRNA, microRNA, or metabolomic profiling, as

well as RNA-Seq and other next generation sequencing experiments. To enable diverse inputs, this application accepts over 20 different types of molecule identifier. As a consequence, this provides users with the potential to combine data from multiple high-throughput sources, channeling them through a single location for concomitant interrogation under identical bioinformatic parameters. This facilitates proteomic data integration with high throughput information spanning multiple regulatory levels, increasing the potential for comprehensive insight into cellular function [43]. A number of user-specified parameters are then applied to define how broad or refined an analysis is desired. For proteomic analysis, protein identities are generally submitted together with expression level data such as fold change, log ratios, or P-values to set prescribed cutoffs for differential expression between experimental groups. Subsets of upregulated and downregulated proteins can then be examined in isolation, or together as a complete differentially expressed cohort. The scope and stringency of input functional relevance is also user controlled, such as breadth of species data to interrogate, whether direct and indirect biological relationships are acceptable, and whether these interactions must be documented relationships only or if predicted interactions are also acceptable.

Pathway analysis output, such as that obtained with IPA and MetaCore, contains enrichment analysis functions highlighting specific functional annotations, canonical pathways, diseases, or other adverse effects overrepresented within the examined dataset (Fig. 14.1). Images representing signaling pathways, metabolic pathways, or other clusters of interest such as protein complexes, may then be opened for inspection, typically with expression data for constituent proteins overlaid on the image for ease of visual representation. Additional predictive elements are now being incorporated to enhance these pathway diagrams. IPA, for example, recently introduced tools designed to predict upstream regulatory effectors with the greatest likelihood of explaining observed input data, including predicted regulator activation states based on observed protein expression values. Moreover, generated pathways overlaid with expression data can now be used to infer whether expression of other known elements within the pathway might also be altered despite not being detected during initial proteomic analysis. Finally, comprehensive pathway analysis algorithms also generate protein interaction and regulatory networks (Fig. 14.1) [4], which can be tailored by settings for maximum network size, number of networks, and whether network nodes are limited to proteins and genes only, or if their composition may be expanded to include other bioactive molecules such as drugs, endogenous chemicals, and metabolites.

What must be kept in mind is that pathway analysis outputs are inferred biological consequences arising from or explaining input proteins and their observed expression values. As these are predictions and not mechanistic explanations, it is best to view pathway analysis as an interpretative tool facilitating hypothesis generation [4]. Ideally, these hypotheses are then tested and validated by experimental follow-up. With the realization that pathway analysis is operating from the standpoint of partial information, and the knowledge that quality and reliability of supported interactions and relationships gleaned from the literature by these applications can be highly variable, the quality of which are not readily apparent without in-depth

analyses of all relevant literature, it is best to approach any results with a healthy dose of skepticism by designing and executing validation experiments whenever possible. Continuous refinement of pathway analysis applications improves as data acquisition increases, but the most convincing systems proteomics studies will always complement predictive analytics with supportive experimental validation.

Pathway analysis algorithms also harbor limitations with respect to generated protein and gene interaction networks [4]. While these networks can be evaluated for enrichment and overrepresentation in a manner similar to that of an unconnected protein or gene dataset that served as initial input, pathway applications are not designed to characterize additional characteristics such as network topology or structure, which imparts additional emergent properties of relevance for particular nodes within the network. Moreover, pathway analysis networks are often intended for visual esthetics rather than functional interpretation, so these applications tend to have upper bounds in their capacity to assemble large networks. As proteomic datasets continue to increase in magnitude, this limitation becomes more problematic for network-oriented biological interpretation. To properly exploit network structure and composition for proteomic systems analysis, it is therefore essential to move beyond pathway network applications and make use of dedicated network analysis tools. Comprehension of some basic principles of complex network analysis, including those that confer value-added properties for systems analysis, facilitates their use for visualizing and interpreting proteomic data.

Complex Network Analysis

What, exactly, is a network? As noted for the hierarchical structure of the complete assemblage of GO terms, a network, or graph, is a collection of nodes, each connecting to one or more additional nodes in a pairwise manner (Table 14.1). In proteomics, then, a network serves as a mathematical representation of known or predicted biological relationships between collections of proteins. Nodes or vertices designate the proteins, while any relationship between two proteins is represented by an edge, or line, connecting the two nodes (Table 14.1). The number of edges connecting a node to other nodes in the network is defined as the first node's degree (Table 14.1). Networks are now understood to assemble into nonrandom structures, where most nodes within the network contain very few connections to other nodes, and thus have a small degree, whereas a much smaller proportion of nodes have many connections, or a large degree.

Once believed to be randomly arranged in terms of connectivity, over the past 15 years this non-stochastic connectivity tendency in biological networks has become better understood. It is now well established that this arrangement of biological network degree distribution (Table 14.1, Fig. 14.1) approximates a power law, leading to a characteristic network topology that is termed scale-free (Table 14.1) [44]. Nodes of extremely high degree are defined as hubs (Table 14.1), and their extensive connectivity is often reflected in these nodes being critical for network functionality.

Another useful network parameter is termed the clustering coefficient (Table 14.1). This is a property of secondary interactions within a network [45], as the clustering coefficient for a particular node indicates the proportion of nodes linking to it that also connect to each other. In other words, this measure defines how interconnected are a node's nearest neighbors. Tightly clustered groups of proteins, which often share similar functional attributes or serve as partners in a multi-protein structural complex, create local regions of high clustering coefficient nodes in a network, and in turn, clusters of clusters can be observed in extremely large networks, such that the network forms a hierarchical structure. Nodes that bridge two or more regions of high clustering within a network are known as bridging nodes (Table 14.1). Due to their position spanning large numbers of nodes on either side, they form a conduit as the shortest path between an inordinately high proportion of node pairs within the network. Therefore, bridging nodes often are also critical to overall network function, like hubs, despite typically being of rather limited degree, unlike hubs. Network non-stochasticity also imparts other emergent properties of biological relevance beyond that of hubs and bridging nodes, such as network structural motifs, modularity, and functional robustness (Table 14.1) [46, 47].

From these rather esoteric descriptions, it may not be readily apparent how networks are useful for representation of proteomic data. Proteins carry out the vast majority of functions within cells, doing so not in isolation but rather in concert with a plethora of other proteins and macromolecules, as components in structural or regulatory interactions, or as part of signaling or metabolic cascades. Accordingly, arrangement of these interactions in the form of a biological network serves as a rational means of assembling complex data in a functional, coherent format. Once generated, biological interaction networks can be evaluated on the basis of their composition via ontological and functional enrichment analysis, and on the basis of network topology or structure, both in terms of its overall architecture as well as by mathematical measures identifying nodes with positions of prominence throughout the network, e.g. hubs and bridging nodes (Table 14.1, Fig. 14.1) [46, 47]. Because proteomic networks are also non-stochastic, regardless of network size or scale, they possess predictable structural characteristics that can be useful for functional interrogation and hypothesis generation. Importantly, such network topology-dependent traits are not readily apparent when their constituent proteins are instead arranged only as lists.

Methods used to generate networks from biological data can differ widely, depending on the source of information used to define interactions, and on underlying presumptions used to evaluate what properties constitute an edge to connect two nodes. When produced in conjunction with pathway analysis, proteomic interaction networks are most often generated using current biological knowledge to establish connectivity. Such networks typically include structural, regulatory, and signaling based interactions, comprising both direct and indirect relationships between nodes. If only a particular subset of interactions is warranted or desired for network assembly, such as protein-protein interactions, these can be assembled directly from the literature or by *de novo* experimental data acquisition. There are also statistically guided methods for network construction, connecting nodes on the basis of co-

expression or correlation [48], or by reverse engineering from expression dynamics using *ab initio* methods [49], although these methods require greater detail regarding data input for modeling and prediction of network interactions than is typically available from proteomic studies. Thus, networks derived from pathway analysis and the accumulated knowledge archived therein is currently the most applicable methodology available for proteomic systems analysis.

Examining protein networks from both a structure and function standpoint requires an understanding of dedicated network analysis and visualization applications. A prominent example is Cytoscape, developed by a multi-institute consortium as an open-source, stand-alone tool for network visualization and evaluation of network structural properties, which over time has added the capacity to access protein and gene data directly from other repositories, thus enhancing network comprehension in biological contexts [50]. As it is open-source, users and developers are welcome to create and contribute new peripherals, or apps, to further advance Cytoscape's utility for network interpretation and interrogation. Newcomers to complex network analysis will appreciate the fact that neither bioinformatic proficiency nor expertise in network biology are required to begin using Cytoscape, although its maturation and broad appeal has led to a plethora of tools and applications that require a substantial commitment to accurately comprehend and exploit to their full potential.

Numerous options exist for visualization of networks constructed in or imported into Cytoscape. Besides the basic requirement for a list of pairwise interactions, additional attributes can be uploaded and appended, such as protein expression data, which can then be superimposed as node or edge attributes or to network layout to organize networks visually on the basis of expression information. Thus, extent and direction of biological change, i.e. up- or down-regulation, can be conveyed by color, shape, or size of nodes and edges, with a variety of optional layouts available to assist researchers in emphasizing particular network elements or properties [11–14]. For example, the layout of nodes in one network can be applied to another in order to co-localize nodes shared by both networks in the same relative spatial regions, enhancing the visual capacity to compare and contrast related networks [14]. Once attributes are applied to network nodes, this information can also be exploited to enable layout co-localization by regional clustering of nodes sharing one or more common attributes [12–14]. Unlike pathway analysis network tools, these functions can be achieved in Cytoscape without concern for upper bounds on network size or structural complexity.

Beyond visualization properties, substantial effort has been devoted to Cytoscape analysis, interrogation and interpretation tools to examine network structure and network functional characteristics. Network Analyzer [51] was developed soon after the introduction of Cytoscape, and was such a popular app for network topology analysis that it is now fully integrated as a standard tool on all new platform downloads. Analyzer assesses a variety of topological elements for both directed and undirected Cytoscape networks, including but not limited to number of nodes and edges, number of connected components, degree distribution, clustering coefficient, average path lengths between pairs of nodes, and network diameter (Table 14.1) [51]. These network topological attributes are not addressed by pathway

analysis network functions, making network specific applications a valuable addition to the systems proteomics repertoire.

From a network functional enrichment standpoint, the most popular Cytoscape app is the Biological Network Gene Ontology (BiNGO) tool [52]. It was designed to acquire current external resources of the GO and apply them internally to fulfill network GO analysis within Cytoscape. Thus, BiNGO interprets biological network over-representation across the breadth of GO domains—biological process, cellular component, and molecular function—by comparison to the entirety of a species-specific reference set, a process similar to enrichment analysis tools in pathway algorithms. BiNGO output can be displayed in two ways. The first is as a significance ranked spreadsheet of terms defined within a particular GO domain, with domain of choice and significance threshold pre-defined in user settings. The second is as a nested hierarchical ontology network representing GO terms as nodes, parent-child relationships as edges, node coloring graded in relation to the presence and extent of statistical significance, and node size scaled to the proportion of initial network nodes mapping to each GO network term. Ultimately, BiNGO generates an ontology network defining the functional attributes of its parent molecular network (Fig. 14.1) [52].

There are now over 250 unique Cytoscape apps designed for specific functions, including but not limited to, import of networks and their attributes, network inference, analysis of existing networks, enrichment and ontology analysis, systems biology, comparison between networks, and communication and scripting applications. Cytoscape has become increasingly popular for cardiovascular proteomic network analysis [11–14, 17, 25] due to the litany of contributors building it into a comprehensive program addressing almost every network-oriented concept imaginable. While described here extensively to outline dedicated network platform applications, Cytoscape is not the only useful network analysis program available, and readers are encouraged to investigate other network visualization and analysis tools. Similar to the Pathguide repository of pathway analysis applications, Graph Visualization Software References formerly served as a single site resource providing information on several dozen network analysis algorithms to enlighten and guide software selection [53], but unfortunately a recent search indicates that this database appears to be no longer available. At this time, no comparable resource is available to guide newcomers to appropriate tools for network analysis, but investigators are encouraged to seek out and apply network associated applications to maximize proteomic systems-oriented data analysis.

Putting the Components Together – Actionable Prognostication with Experimental Validation

Unlike traditional reductionist approaches where hypotheses are formulated and subsequently investigated by applying various molecular biology techniques, often with an emphasis on characterizing function of only a single protein or biological pathway, proteomic and other high throughput techniques are often applied without

a preconceived notion of what may or will be detected or discovered. Indeed, the biology may not even be sufficiently well understood to formulate actionable hypotheses until after such data is first analyzed and interpreted. Accordingly, high throughput analyses are often viewed or approached from a different scientific standpoint, wherein data analysis serves as the hypothesis generating step that must subsequently be validated (Fig. 14.1). Thus, acquired proteomic data does not typically serve as a final answer in and of itself. Instead, delivering on the promise of proteomic data often requires the application of actionable prognostication.

A case in point is proteomic comprehension of the cardiac implications of ATP-sensitive K^+ (K_{ATP}) channel deficiency, caused by absence of the *KCNJ11*-encoded Kir6.2 pore forming subunit of the channel multi-subunit protein complex [54]. Functional consequence of genetic knockout in cardiac myocytes is a loss of K^+ conductance across the cell membrane, but K_{ATP} channel activity influences far more, modulating membrane potential-dependent cellular metabolism much like a rheostat, adjusting function to match cellular energy demands [55–58]. Even though K_{ATP} channelopathies are implicated in human cardiac disease [59–61], consequences of channel deficiency predisposing to disease vulnerability escaped broader molecular comprehension, mandating proteomic systems interrogation of channel dysfunction in various contexts [12–15]. In the *KCNJ11*-knockout, rather than simply being a case of presence *versus* absence of a single protein, proteomic analysis determined that, even in the absence of superimposed cardiac stress, more than 100 proteins were significantly altered in response to chronic K_{ATP} channel deficiency [13]. Taking this list of proteins through a network systems analysis is particularly revealing for comprehension of the underlying mechanistic consequences of channel dysfunction.

Initial ontological classification (**Step 1 of what to do with your proteomic list**) indicated that a little over 60 % of differentially expressed proteins could be categorized as having direct involvement in metabolic function, whereas the remainder participated in a variety of other cellular processes, including proteolysis, chaperones, cytostructure, oxidoreductases, transcription or translation, and regulation of cell signaling [13]. This abundance of metabolic connections is consistent with the known impact of the channel as a metabolic rheostat, the prominence of which was reinforced by IPA functional ontology classification (**Step 2 of what to do with your proteomic list**) [13]. Moreover, metabolic relevance was further strengthened by subsequent complex network ontology enrichment analysis (**Step 2 applied to output acquired from Step 3 of what to do with your proteomic list**). In this regard, BiNGO analysis was conducted within Cytoscape to define overrepresented biological processes – one of the three primary GO domains – associated with the expanded K_{ATP} channel–dependent metaboproteome (Table 14.1) network derived from the remodeled proteome. The resultant BiNGO ontology network comprised nearly 1,000 distinct GO terms associating with the parental molecular network, yet only 55 of these were significantly overrepresented [13]. Moreover, every one of the 55 was a metabolic process, collectively forming a highly nested ontology network within a limited number of broader metabolic functions. Primarily overrepresented were GO terms involved in glycolysis, as well as tricarboxylic acid cycle, fatty acid,

and other substrate metabolism branches, along with some degree of protein catabolism enrichment [13].

These parameters provide a sense of altered proteome functional attributes, but not of the biological consequences of proteome remodeling. Pathway analysis was thus also applied in a complementary manner to predict potential adverse effects, yielding actionable insight into the implications arising from and consistent with the altered proteome. Here, "cardiovascular disease" was significantly overrepresented at the level of the proteome and, even more extensively, at the interactome (Table 14.1) level integrating all proteome changes in their broader network neighborhood [13]. Experimental evidence supporting susceptibility of the K_{ATP} channel deficient cohort to cardiovascular disease was evident in measures of cardiac mass, cardiac function, and survivorship in response to increasing levels of imposed cardiac stress [13]. Thus, proteomic network systems analysis here incorporated actionable experimental evaluation, validating the predicted functional consequences of observed proteome remodeling (**Step 4, actionable prognostication with experimental validation**).

Similar systems approaches have also been applied to understand proteomic consequences of K_{ATP} channel deficiency in the setting of superimposed cardiac stress [12, 14]. For example, prediction of overrepresented adverse effects facilitated experimental follow-up in a model of deoxycorticosteroid and salt-induced hypertension, where pathway analysis of proteomic data predicted three adverse effects related exclusively to cardiac function – cardiac damage, cardiac enlargement, and cardiac fibrosis [12]. Each effect was subsequently confirmed by assessment of cardiac output, measurement of heart-to-body-weight ratios, and evaluation of collagen deposition, respectively, validating predicted detrimental cardiac effects of K_{ATP} channel-dependent proteome remodeling in response to physiological stress [12]. Pathway analysis adverse effect screening also proved instrumental in evaluating consequences of proteome remodeling in cardiomyopathy and the structural and functional remodeling mediated by the response to stem cell therapy in cardiomyopathic hearts [14]. Here, proteome changes associated with cardiomyopathy were subjected to pathway analysis, with *in silico* prediction of both enrichment of cardiac disease as well as several cardiac adverse effects, which were greatly ameliorated or completely absent when evaluating the stem cell treated cardiomyopathic proteome [14]. Prognostication was validated by a range of echocardiographic metrics and anatomical measurements, confirming predicted deleterious structural and functional outcomes of disease and their improvement following cell mediated therapy [14].

Further supporting network approaches, a distinct benefit of extending systems analysis to complex networks is their potential to provide value-added elements to evaluate implications of the proteomic data in a network-oriented context. Network structure, i.e. topology, can be exploited on the basis of identification and targeted inhibition of network hubs. While most nodes in scale-free networks possess a low degree and can be removed or inhibited without great risk of leading to a loss in network functionality, i.e. robustness, the reverse is also true, wherein targeting highly connected nodes can be exploited to evaluate whether a predicted network

function or other emergent property is dependent on network integrity mediated by their hubs [62]. For instance, such connectivity properties suggest that inhibition of one or more primary hubs of an endodermal secretome network might prevent its potentiating effect on cardiac differentiation [11]. This prediction was reinforced by *in silico* modulation of composite network generation, whereby prioritization of cardiovascular development predicted for the network was demoted after arbitrarily removing the most highly connected hub from pathway analysis input data [11]. Indeed, this was demonstrated functionally when application of pharmacological inhibitors of the two most highly connected nodes each abolished the cardiac potentiation effect mediated by the secretome. This included the primary hub that was detected during initial proteomic analysis as well as the secondary hub that was only incorporated during network generation but was noticeably absent from the proteomic data [11]. Thus, network topology assessment has the potential to provide value-added emergent properties for hypothesis generation of intra-network positional relevance. Moreover, even though network generation increases overall molecular complexity by adding more proteins to the initial proteomic list, in doing so it also yields further potentially relevant candidates for systems evaluation that may be critical contributors to the underlying biology that were nevertheless overlooked during initial proteomic screening [4, 5, 11].

Conclusion

Continued technological advances, with improved instrument sensitivity and resolution combined with expanded, more detailed protein databases, will lead to increasingly larger proteomic datasets, each harboring tremendous biological intricacy. Network systems analysis strategies will therefore become ever more critical for proteomic and other high throughput data deconvolution [4, 63]. Herein, guidance is provided on generalized analytic approaches by which to systematically organize, cluster, and prioritize proteomic datasets in their entirety to reduce information complexity while simultaneously yielding functional insights. An important qualifier is that ontological classification, and enrichment, pathway, and complex network analyses may be considered as interchangeable modular components of a network systems approach. Rather than applying each as a stand-alone topic or steps that must be adhered to in a prescribed order, these elements may be arranged flexibly and used in a variety of ways, as required to address a specific biological question. Indeed, the same step may even be repeated multiple times at different stages in an analysis. For instance, enrichment analysis can be conducted on initial ontological categories, during pathway analysis, or on final output networks [6], potentially revealing shifts in focus at successive points in an analysis. When actionable hypotheses are generated via systems analysis and examined experimentally, refinement or modification of the initial interpretation may potentially mandate an additional round of high throughput data acquisition. Thus, network systems analysis can also be viewed as an iterative process, with a

cyclical transition from high throughput data to interrogation, followed by experimentation to validate or refine interpretation, ultimately guiding subsequent decisions on additional proteomic or other high throughput experiments (Fig. 14.1) [5, 6]. Comprehension of the basis and utility of these organizing and interpretive principles provides a foundation for their application, preparing students, proteomic practitioners, and clinicians alike for effective application of basic and translational proteomic network medicine to further our understanding of cardiovascular health and disease.

References

1. Agnetti G, Lindsey ML, Foster DB, editors. Manual of cardiovascular proteomics. Cham, Switzerland: Springer; 2016.
2. Oswald ME, Grosjean S. Confirmation Bias. In: Pohl RF, editor. Cognitive illusions: a handbook on fallacies and biases in thinking, judgement and memory. Hove, UK, and New York, USA: Psychology Press; 2004. p. 79–96.
3. Jeng M. A selected history of expectation bias in physics. Am J Phys. 2006;74:578–83.
4. Arrell DK, Terzic A. Systems proteomics for translational network medicine. Circ Cardiovasc Genet. 2012;5:o8–16.
5. Arrell DK, Terzic A. Network systems biology for drug discovery. Clin Pharmacol Ther. 2010;88:120–5.
6. Lindsey ML, Mayr M, Gomes AV, Delles C, Arrell DK, Murphy AM, Lange RA, Costello CE, Jin YF, Laskowitz DT, Sam F, Terzic A, Van Eyk J, Srinivas PR. Transformative impact of proteomics on cardiovascular health and disease: a scientific statement from the American Heart Association. Circulation. 2015;132:852–72.
7. Levin Y. The role of statistical power analysis in quantitative proteomics. Proteomics. 2011;11:2565–7.
8. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G. Gene Ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet. 2000;25:25–9.
9. Pundir S, Magrane M, Martin MJ, O'Donovan C; UniProt Consortium. Searching and navigating UniProt databases. Curr Protoc Bioinformatics. 2015;50:1.27.1–1.27.10.
10. Huntley RP, Sawford T, Mutowo-Meullenet P, Shypitsyna A, Bonilla C, Martin MJ, O'Donovan C. The GOA database: Gene Ontology annotation updates for 2015. Nucleic Acids Res. 2015;43(Database issue):D1057–63.
11. Arrell DK, Niederländer NJ, Faustino RS, Behfar A, Terzic A. Cardioinductive network guiding stem cell differentiation revealed by proteomic cartography of tumor necrosis factor alpha-primed endodermal secretome. Stem Cells. 2008;26:387–400.
12. Zlatkovic J, Arrell DK, Kane GC, Miki T, Seino S, Terzic A. Proteomic profiling of K_{ATP} channel-deficient hypertensive heart maps risk for maladaptive cardiomyopathic outcome. Proteomics. 2009;9:1314–25.

13. Arrell DK, Zlatkovic J, Kane GC, Yamada S, Terzic A. ATP-sensitive K$^+$ channel knockout induces cardiac proteome remodeling predictive of heart disease susceptibility. J Proteome Res. 2009;8:4823–34.
14. Zlatkovic-Lindor J, Arrell DK, Yamada S, Nelson TJ, Terzic A. ATP-sensitive K$^+$ channel-deficient dilated cardiomyopathy proteome remodeled by embryonic stem cell therapy. Stem Cells. 2010;28:1355–67.
15. Arrell DK, Zlatkovic-Lindor J, Yamada S, Terzic A. K$_{ATP}$ channel-dependent metaboproteome decoded: systems approaches to heart failure prediction, diagnosis, and therapy. Cardiovasc Res. 2011;90:258–66.
16. Foster DB, Liu T, Rucker J, O'Meally RN, Devine LR, Cole RN, O'Rourke B. The cardiac acetyl-lysine proteome. PLoS One. 2013;8:e67513.
17. Isserlin R, Merico D, Alikhani-Koupaei R, Gramolini A, Bader GD, Emili A. Pathway analysis of dilated cardiomyopathy using global proteomic profiling and enrichment maps. Proteomics. 2010;10:1316–27.
18. Monte E, Mouillesseaux K, Chen H, Kimball T, Ren S, Wang Y, Chen JN, Vondriska TM, Franklin S. Systems proteomics of cardiac chromatin identifies nucleolin as a regulator of growth and cellular plasticity in cardiomyocytes. Am J Physiol Heart Circ Physiol. 2013;305:H1624–38.
19. Drozdov I, Didangelos A, Yin X, Zampetaki A, Abonnenc M, Murdoch C, Zhang M, Ouzounis CA, Mayr M, Tsoka S, Shah AM. Gene network and proteomic analyses of cardiac responses to pathological and physiological stress. Circ Cardiovasc Genet. 2013;6:588–97.
20. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A. 2005;102:15545–50.
21. Cha S, Imielinski MB, Rejtar T, Richardson EA, Thakur D, Sgroi DC, Karger BL. In situ proteomic analysis of human breast cancer epithelial cells using laser capture microdissection: annotation by protein set enrichment analysis and gene ontology. Mol Cell Proteomics. 2010;9:2529–44.
22. Lavallée-Adam M, Rauniyar N, McClatchy DB, Yates JR 3rd. PSEA-Quant: a protein set enrichment analysis on label-free and label-based protein quantification data. J Proteome Res. 2014;13:5496–509.
23. Bader GD, Cary MP, Sander C. Pathguide: a pathway resource list. Nucleic Acids Res. 2006;34(Database issue):D504–6.
24. Demir E, Cary MP, Paley S, Fukuda K, Lemer C, Vastrik I, Wu G, D'Eustachio P, Schaefer C, Luciano J, Schacherer F, Martinez-Flores I, Hu Z, Jimenez-Jacinto V, Joshi-Tope G, Kandasamy K, Lopez-Fuentes AC, Mi H, Pichler E, Rodchenkov I, Splendiani A, Tkachev S, Zucker J, Gopinath G, Rajasimha H, Ramakrishnan R, Shah I, Syed M, Anwar N, Babur O, Blinov M, Brauner E, Corwin D, Donaldson S, Gibbons F, Goldberg R, Hornbeck P, Luna A, Murray-Rust P, Neumann E, Ruebenacker O, Samwald M, van Iersel M, Wimalaratne S, Allen K, Braun B, Whirl-Carrillo M, Cheung KH, Dahlquist K, Finney A, Gillespie M, Glass E, Gong L, Haw R, Honig M, Hubaut O, Kane D, Krupa S, Kutmon M, Leonard J, Marks D, Merberg D, Petri V, Pico A, Ravenscroft D, Ren L, Shah N, Sunshine M, Tang R, Whaley R, Letovksy S, Buetow KH, Rzhetsky A, Schachter V, Sobral BS, Dogrusoz U, McWeeney S, Aladjem M, Birney E, Collado-Vides J, Goto S, Hucka M, Le Novère N, Maltsev N, Pandey A, Thomas P, Wingender E, Karp PD, Sander C, Bader GD. The BioPAX community standard for pathway data sharing. Nat Biotechnol. 2010;28:935–42.
25. Behfar A, Perez-Terzic C, Faustino RS, Arrell DK, Hodgson DM, Yamada S, Puceat M, Niederländer N, Alekseev AE, Zingman LV, Terzic A. Cardiopoietic programming of embryonic stem cells for tumor-free heart repair. J Exp Med. 2007;204:405–20.
26. Fu Z, Wang M, Gucek M, Zhang J, Wu J, Jiang L, Monticone RE, Khazan B, Telljohann R, Mattison J, Sheng S, Cole RN, Spinetti G, Pintus G, Liu L, Kolodgie FD, Virmani R, Spurgeon H, Ingram DK, Everett AD, Lakatta EG, Van Eyk JE. Milk fat globule protein epidermal

growth factor-8: a pivotal relay element within the angiotensin II and monocyte chemoattractant protein-1 signaling cascade mediating vascular smooth muscle cells invasion. Circ Res. 2009;104:1337–46.

27. Mayr M, Zampetaki A, Sidibe A, Mayr U, Yin X, De Souza AI, Chung YL, Madhu B, Quax PH, Hu Y, Griffiths JR, Xu Q. Proteomic and metabolomic analysis of smooth muscle cells derived from the arterial media and adventitial progenitors of apolipoprotein E-deficient mice. Circ Res. 2008;102:1046–56.

28. Grant JE, Bradshaw AD, Schwacke JH, Baicu CF, Zile MR, Schey KL. Quantification of protein expression changes in the aging left ventricle of Rattus norvegicus. J Proteome Res. 2009;8:4252–63.

29. Li Z, Liu N, Zhang LS, Gong K, Cai Y, Gao W, Liu Z, Liu S, Han Q, Zhang Y. Proteomic profiling reveals comprehensive insights into adrenergic receptor-mediated hypertrophy in neonatal rat cardiomyocytes. Proteomics Clin Appl. 2009;3:1407–21.

30. Mayr M, Liem D, Zhang J, Li X, Avliyakulov NK, Yang JI, Young G, Vondriska TM, Ladroue C, Madhu B, Griffiths JR, Gomes A, Xu Q, Ping P. Proteomic and metabolomic analysis of cardioprotection: interplay between protein kinase C epsilon and delta in regulating glucose metabolism of murine hearts. J Mol Cell Cardiol. 2009;46:268–77.

31. Agnetti G, Kaludercic N, Kane LA, Elliott ST, Guo Y, Chakir K, Samantapudi D, Paolocci N, Tomaselli GF, Kass DA, Van Eyk JE. Modulation of mitochondrial proteome and improved mitochondrial function by biventricular pacing of dyssynchronous failing hearts. Circ Cardiovasc Genet. 2010;3:78–87.

32. Bugger H, Schwarzer M, Chen D, Schrepper A, Amorim PA, Schoepe M, Nguyen TD, Mohr FW, Khalimonchuk O, Weimer BC, Doenst T. Proteomic remodelling of mitochondrial oxidative pathways in pressure overload-induced heart failure. Cardiovasc Res. 2010;85:376–84.

33. Farina A, Chambery A, Esposito S, Agozzino L, Cotrufo M, Della Corte A, Parente A. Proteomic profiling of medial degeneration in human ascending aorta. Clin Biochem. 2010;43:387–96.

34. Parguiña AF, Fernández Parguiña A, Grigorian-Shamajian L, Agra RM, Teijeira-Fernández E, Rosa I, Alonso J, Viñuela-Roldán JE, Seoane A, González-Juanatey JR, García A. Proteins involved in platelet signaling are differentially regulated in acute coronary syndrome: a proteomic study. PLoS One. 2010;5:e13404.

35. Angel PM, Nusinow D, Brown CB, Violette K, Barnett JV, Zhang B, Baldwin HS, Caprioli RM. Networked-based characterization of extracellular matrix proteins from adult mouse pulmonary and aortic valves. J Proteome Res. 2011;10:812–23.

36. Fogle RL, Hollenbeak CS, Stanley BA, Vary TC, Kimball SR, Lynch CJ. Functional proteomic analysis reveals sex-dependent differences in structural and energy-producing myocardial proteins in rat model of alcoholic cardiomyopathy. Physiol Genomics. 2011;43:346–56.

37. Hammer E, Goritzka M, Ameling S, Darm K, Steil L, Klingel K, Trimpert C, Herda LR, Dörr M, Kroemer HK, Kandolf R, Staudt A, Felix SB, Völker U. Characterization of the human myocardial proteome in inflammatory dilated cardiomyopathy by label-free quantitative shotgun proteomics of heart biopsies. J Proteome Res. 2011;10:2161–71.

38. Hanzu FA, Palomo M, Kalko SG, Parrizas M, Garaulet M, Escolar G, Gomis R, Diaz-Ricart M. Translational evidence of endothelial damage in obese individuals: inflammatory and prothrombotic responses. J Thromb Haemost. 2011;9:1236–45.

39. Qi YX, Jiang J, Jiang XH, Wang XD, Ji SY, Han Y, Long DK, Shen BR, Yan ZQ, Chien S, Jiang ZL. PDGF-BB and TGF-β1 on cross-talk between endothelial and smooth muscle cells in vascular remodeling induced by low shear stress. Proc Natl Acad Sci U S A. 2011;108: 1908–13.

40. Li W, Rong R, Zhao S, Zhu X, Zhang K, Xiong X, Yu X, Cui Q, Li S, Chen L, Cai J, Du J. Proteomic analysis of metabolic, cytoskeletal and stress response proteins in human heart failure. J Cell Mol Med. 2012;16:59–71.

41. de la Cuesta F, Barderas MG, Calvo E, Zubiri I, Maroto AS, Darde VM, Martin-Rojas T, Gil-Dones F, Posada-Ayala M, Tejerina T, Lopez JA, Vivanco F, Alvarez-Llamas G. Secretome

analysis of atherosclerotic and non-atherosclerotic arteries reveals dynamic extracellular remodeling during pathogenesis. J Proteomics. 2012;75:2960–71.

42. Sun Z, Biela LM, Hamilton KL, Reardon KF. Concentration-dependent effects of the soy phytoestrogen genistein on the proteome of cultured cardiomyocytes. J Proteomics. 2012;75:3592–604.

43. Lu R, Markowetz F, Unwin RD, Leek JT, Airoldi EM, MacArthur BD, Lachmann A, Rozov R, Ma'ayan A, Boyer LA, Troyanskaya OG, Whetton AD, Lemischka IR. Systems-level dynamic analyses of fate change in murine embryonic stem cells. Nature. 2009;462:358–62.

44. Barabási A-L, Albert R. Emergence of scaling in random networks. Science. 1999;286:509–12.

45. Arrell DK, Terzic A. Interpreting networks in systems biology. Clin Pharmacol Ther. 2013;93:389–92.

46. Barabási A-L, Oltvai ZN. Network biology: understanding the cell's functional organization. Nat Rev Genet. 2004;5:101–13.

47. Newman MEJ. The structure and function of complex networks. SIAM Rev. 2003;45:167–256.

48. Dewey FE, Perez MV, Wheeler MT, Watt C, Spin J, Langfelder P, Horvath S, Hannenhalli S, Cappola TP, Ashley EA. Gene coexpression network topology of cardiac development, hypertrophy, and failure. Circ Cardiovasc Genet. 2011;4:26–35.

49. Khalil I, Brewer MA, Neyarapally T, Runowicz CD. The potential of biologic network models in understanding the etiopathogenesis of ovarian cancer. Gynecol Oncol. 2010;116:282–5.

50. Cline MS, Smoot M, Cerami E, Kuchinsky A, Landys N, Workman C, Christmas R, Avila-Campilo I, Creech M, Gross B, Hanspers K, Isserlin R, Kelley R, Killcoyne S, Lotia S, Maere S, Morris J, Ono K, Pavlovic V, Pico AR, Vailaya A, Wang PL, Adler A, Conklin BR, Hood L, Kuiper M, Sander C, Schmulevich I, Schwikowski B, Warner GJ, Ideker T, Bader GD. Integration of biological networks and gene expression data using Cytoscape. Nat Protoc. 2007;2:2366–82.

51. Assenov Y, Ramírez F, Schelhorn SE, Lengauer T, Albrecht M. Computing topological parameters of biological networks. Bioinformatics. 2008;24:282–4.

52. Maere S, Heymans K, Kuiper M. BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. Bioinformatics. 2005;21:3448–9.

53. Pinaud B, Kuntz P. GVSR: an on-line guide for choosing a graph visualization software. In: Brandes U, Cornelsen S, editors. Graph drawing: 18th international symposium, GD 2010, lecture notes in computer science. Berlin: Springer; 2011. p. 400–1.

54. Miki T, Nagashima K, Tashiro F, Kotake K, Yoshitomi H, Tamamoto A, Gonoi T, Iwanaga T, Miyazaki J, Seino S. Defective insulin secretion and enhanced insulin action in K_{ATP} channel-deficient mice. Proc Natl Acad Sci U S A. 1998;95:10402–6.

55. Zingman LV, Alekseev AE, Bienengraeber M, Hodgson DM, Karger AB, Dzeja PP, Terzic A. Signaling in channel/enzyme multimers: ATPase transitions in SUR module gate ATP-sensitive K^+ conductance. Neuron. 2001;31:233–45.

56. Liu XK, Yamada S, Kane GC, Alekseev AE, Hodgson DM, O'Cochlain DF, Jahangir A, Miki T, Seino S, Terzic A. Genetic disruption of Kir6.2, the pore-forming subunit of ATP-sensitive K^+ channel, predisposes to catecholamine induced ventricular dysrhythmia. Diabetes. 2004;53:S165–8.

57. Selivanov VA, Alekseev AE, Hodgson DM, Dzeja PP, Terzic A. Nucleotide-gated K_{ATP} channels integrated with creatine and adenylate kinases: amplification, tuning and sensing of energetic signals in the compartmentalized cellular environment. Mol Cell Biochem. 2004;256–7:243–56.

58. Gumina RJ, O'Cochlain DF, Kurtz CE, Bast PH, Pucar D, Mishra P, Miki T, Seino S, Macura S, Terzic A. K_{ATP} channel knockout worsens myocardial calcium stress load in vivo and impairs recovery in stunned heart. Am J Physiol Heart Circ Physiol. 2007;292:H1706–13.

59. Yamada S, Kane GC, Behfar A, Liu XK, Dyer RB, Faustino RS, Miki T, Seino S, Terzic A. Protection conferred by myocardial ATP-sensitive K^+ channels in pressure overload induced

congestive heart failure revealed in KCNJ11 Kir6.2-null mutant. J Physiol. 2006;577: 1053–65.

60. Bienengraeber M, Olson TM, Selivanov VA, Kathmann EC, O'Cochlain DF, Gao F, Karger AB, Ballew JD, Hodgson DM, Zingman LV, Pang Y-P, Alekseev AE, Terzic A. ABCC9 mutations identified in human dilated cardiomyopathy disrupt catalytic K_{ATP} channel gating. Nat Genet. 2004;36:382–7.

61. Olson TM, Alekseev AE, Moreau C, Liu XK, Zingman LV, Miki T, Seino S, Asirvatham SJ, Jahangir A, Terzic A. K_{ATP} channel mutation confers risk for vein of Marshall adrenergic atrial fibrillation. Nat Clin Pract Cardiovasc Med. 2007;4:110–6.

62. Albert R, Jeong H, Barabási A-L. Error and attack tolerance of complex networks. Nature. 2000;406:378–82.

63. Loscalzo J, Barabási A-L. Systems biology and the future of medicine. Wiley Interdiscip Rev Syst Biol Med. 2011;3:619–27.