

Review

Network-Based Methods for Approaching Human Pathologies from a Phenotypic Point of View

Juan A. G. Ranea ^{1,2,3,4,†}, James Perkins ^{1,2,3,4,†}, Mónica Chagoyen ⁵, Elena Díaz-Santiago ¹ and Florencio Pazos ^{5,*}

¹ Department of Molecular Biology and Biochemistry, University of Malaga, 29071 Malaga, Spain; ranea@uma.es (J.A.G.R.); jimrperkins@gmail.com (J.P.); elenadiasantiago.92@gmail.com (E.D.-S.)

² CIBER de Enfermedades Raras, Instituto de Salud Carlos III, 28029 Madrid, Spain

³ Institute of Biomedical Research in Malaga (IBIMA), 29071 Malaga, Spain

⁴ Spanish National Bioinformatics Institute (INB/ELIXIR-ES), Instituto de Salud Carlos III (ISCIII), 28020 Madrid, Spain

⁵ Computational Systems Biology Group, Systems Biology Department, National Centre for Biotechnology (CNB-CSIC), 28049 Madrid, Spain; monica.chagoyen@cnb.csic.es

* Correspondence: pazos@cnb.csic.es

† These authors contributed equally to this work.

Abstract: Network and systemic approaches to studying human pathologies are helping us to gain insight into the molecular mechanisms of and potential therapeutic interventions for human diseases, especially for complex diseases where large numbers of genes are involved. The complex human pathological landscape is traditionally partitioned into discrete “diseases”; however, that partition is sometimes problematic, as diseases are highly heterogeneous and can differ greatly from one patient to another. Moreover, for many pathological states, the set of symptoms (phenotypes) manifested by the patient is not enough to diagnose a particular disease. On the contrary, phenotypes, by definition, are directly observable and can be closer to the molecular basis of the pathology. These clinical phenotypes are also important for personalised medicine, as they can help stratify patients and design personalised interventions. For these reasons, network and systemic approaches to pathologies are gradually incorporating phenotypic information. This review covers the current landscape of phenotype-centred network approaches to study different aspects of human diseases.

Keywords: biological network; network medicine; pathological phenotype; gene prioritization



Citation: Ranea, J.A.G.; Perkins, J.; Chagoyen, M.; Díaz-Santiago, E.; Pazos, F. Network-Based Methods for Approaching Human Pathologies from a Phenotypic Point of View. *Genes* **2022**, *13*, 1081. <https://doi.org/10.3390/genes13061081>

Academic Editors: Stefania Bortoluzzi, Piero Fariselli and Anelia D. Horvath

Received: 13 May 2022

Accepted: 14 June 2022

Published: 17 June 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Living systems are characterised by a large number of components immersed in intricate networks of interactions, making them prototypical examples of complex systems. As such, many of their properties cannot be understood through the reductionist approach of molecular biology, which is based on the detailed characterization of individual molecular components, under the assumption that the properties of the system can be obtained from a simple combination of these. This approach fails to adequately reflect many aspects of living systems, which cannot be fully explained by a simple (additive) combination of their constituent molecular components [1–4]. In many cases, the properties of these components do not mean much outside the molecular level itself, and it is only in the context of the complex network of interactions and relationships with other components they acquire a biological meaning.

An alternative approach to biological phenomena from a systemic point of view, typically referred to as Systems Biology [5,6], complements the reductionist approach of Molecular Biology by studying living systems not from the point of view of their molecular components (“bottom-up”), but from higher levels of biological complexity (“top-down”). Frequently, these higher levels are represented by large and complex networks coding relationships between molecular components. Molecular networks are representations of

relationships between different molecular components, with “relationship” often being defined in a broad sense. The discipline that mines these networks to extract biological knowledge is Network Biology [7,8]. Network Biology applies the tools of network analysis and graph theory to extract useful information from large molecular networks.

Although these systemic approaches are not new, they are experiencing a resurgence due to the ongoing revolution in “-omics” techniques, which allow one to obtain massive sets of relationships between molecular components that can be used to build molecular networks.

These systemic and network methodologies are not only being applied to basic research but to biotechnological and biomedical problems as well. Many pathologies cannot be reduced to a failure in a single gene or a small number of genes in a simple, additive way. These complex diseases are better reflected at the “network level”, allowing the integration of information on the relationships between genes, drugs, environmental factors and more. The discipline that approaches human pathologies from this systemic point of view is often called Network Medicine [9–12].

In medicine in general, and network medicine in particular, the basic unit for partitioning the complex human pathological landscape into discrete entities has been the “disease”. Nevertheless, this continuous landscape can also be partitioned into clinical signs, also known as symptoms or pathological phenotypes, which can be defined as the external manifestations of a pathological state (disease). These include entities such as fever, haemorrhage, inflammation, seizures, etc. Two unrelated diseases can share common phenotypes; conversely, two related diseases, or even the same disease manifested in different individuals, can show different sets of phenotypes. Although diseases were the basic units in the first network-based approaches to pathologies, network-medicine approaches focused on clinical phenotypes are becoming increasingly popular.

In this review, we provide an overview of these phenotype-centred network approaches to human diseases. We start with a short introduction to the main network concepts and general approaches used in network medicine, and continue with examples of applications to phenotypes. We also include a table with a summary of the main resources related this subject.

2. Overview of Network Approaches for Studying Human Pathologies

A network (“Graph”, in mathematical terms) is just a representation of relationships between entities. Any phenomenon that can be represented as entities linked by relationships can be modelled as a graph. The entities are usually called nodes or vertices, and the relationships edges (Figure 1a). Both nodes and edges represent entities and relationships understood in the broadest possible sense. A node can represent a physical entity, such as a protein, gene, person, or even a computer, or other non-physical concepts, such as cell state, developmental stage, disease, or computer software. Similarly, edges can represent any type of generic relationship between nodes, such as physical interactions between proteins, chemical transformations between metabolites, hypertext links between two web pages, or subroutine calls in a computer program. Networks with more than one type of node or edge (e.g., some nodes representing proteins and other metabolites) are termed multipartite networks (e.g., Figure 2).

Network Medicine is largely concerned with the study of molecular networks, which represent generic relationships between molecular entities (Figure 1a). In general, these approaches try to extract disease-related information from the complex topological patterns present in these large networks.

A plethora of methodologies have been devised for extracting disease-related information from the topology of a molecular network. For more comprehensive reviews see [8,9,11]. A key concept in most of these approaches is that of the “disease-related module”. A module (also known as “cluster”) is a purely topological concept, and describes a set of network nodes that are enriched in internal connections (connections between themselves) compared to external connections (connections with other nodes of the net-

work) (Figure 1). Topologically, they represent sub-networks that present some degree of independence from the rest of the network. In biological networks, these topological modules have been shown to represent functional modules, comprising functionally related sets of molecular entities [13]. For example, in protein interaction networks, it has been shown that they correspond to interacting proteins involved in the same biological process and/or forming a molecular complex (e.g., ribosome, proteasome) or working together in a signalling pathway [14–16]; in metabolic networks, they typically correspond to related metabolites involved in the same metabolic pathway [17]; in gene regulatory networks, they correspond to sets of related transcription factors involved in controlling a given cellular process [18]. Consequently, topological modules of biological networks can be considered functional modules (Figure 1b). Interestingly, it has also been shown that topological modules often represent disease-related modules, in the sense that genes associated with a given disease, when mapped into a biological network, tend to cluster together (e.g., [19–23]) (Figure 1b). In cancer, a complex disease characterised by a progressive accumulation of mutations, it has been shown that it is the concentration of mutated genes in network modules that characterises the transition from health to disease, rather than the general increase in the number of mutations [20]. Even in very complex diseases involving hundreds to thousands of genes, these tend to concentrate in a reduced number of modules/pathways (e.g., autism [24]). This relationship of topological modules with functional and disease-related modules forms the basis of most approaches in Network Medicine. This allows us to connect diseases with their underlying molecular mechanisms.

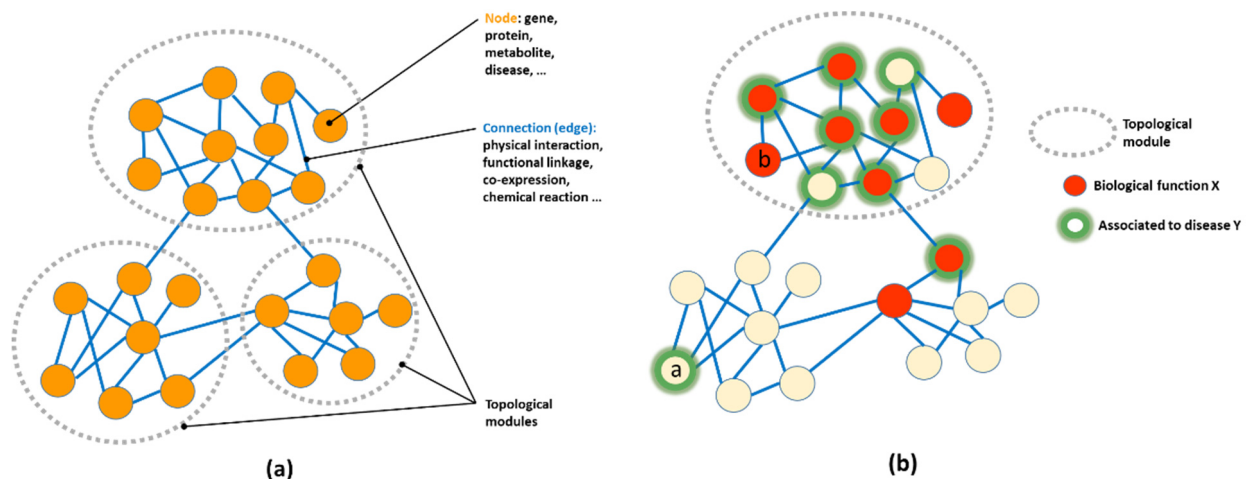


Figure 1. (a) Schematic representation of a generic biological network coding linkages between different molecular entities. This schematic network has three topological modules (clusters). Real biological networks can have tens of thousands of nodes and hundreds of thousands of relationships. (b) Relationships between topological, functional, and disease-related modules. The proteins involved in a specific function (“X”) are coloured red. Those associated with a given disease (e.g., those whose mutation is known to cause “disease Y”) are highlighted with green halos. Proteins known to be involved in “X” tend to cluster in the network. Those associated with disease “Y” tend to cluster in the same topological module, indicating that disease “Y” may be related to a malfunction of the biological process “X”. Alterations of other proteins in the same topological/functional module may also lead to the same disease as they disrupt the same process, even if they have not been detected yet (e.g., gene “b”). Conversely, genes far apart from the cluster might be discarded (e.g., “a”). Network propagation methods would tend to expand that initial set of Y-associated genes to the whole topological cluster as well as discard nonrelated genes.

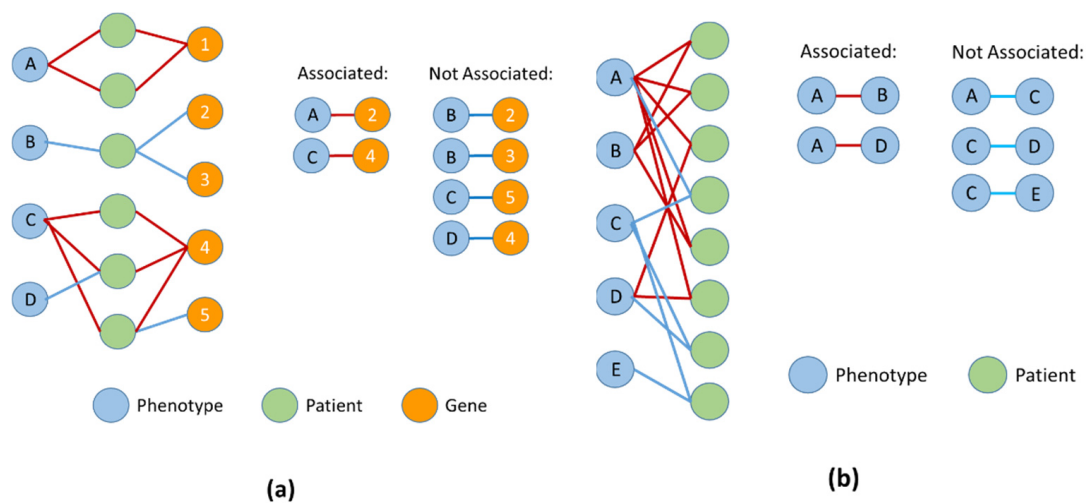


Figure 2. Overview of the use of multipartite networks to obtain phenotype–gene and phenotype–phenotype associations. **(a)** A phenotype–patient–gene tripartite network is constructed from patient data, such that a given phenotype in the cohort is linked to a gene in the cohort when there is at least one patient in the cohort who both manifests the phenotype and has a mutation that maps to the gene. Once the network has been built, it can be analysed to find those phenotype–gene pairs that are highly and specifically connected over many patients, using a statistical test to determine whether there is evidence that the pair is associated. There is also a third set of phenotype gene pairs, which are not connected via any patient (e.g., A-2), these are also considered not to be associated. **(b)** A phenotype–patient bipartite network is constructed from patient data, such that phenotypes are connected to patients when there is at least one patient in the cohort manifesting the phenotype. Once built, the network can be analysed to find pairs of associated phenotypes that are connected by many patients in a specific manner, using a statistical test. As with the phenotype–gene pairs, there are phenotype–phenotype pairs that are not connected by any patient (e.g., B–C). These must also be considered not associated.

There are many approaches for locating these disease-related modules from an initial set of “seed” genes/proteins associated with a given disease [25,26]. These methods, usually termed “network propagation” or “network diffusion” approaches [26,27], detect the topological modules enriched in these seed genes (Figure 1b) using a variety of strategies. The seed genes are those known to be associated with the disease according to different pieces of evidence, such as phenotypic (e.g., disease-associated expression change) or genotypic data (mutations in affected individuals) [28].

Locating these disease-related modules is important for different reasons. On one hand, it allows the filtering of the original set of genes, discarding or adding new ones based on their belonging/closeness to the module (e.g., genes “a” and “b” respectively in Figure 1b). This is related to approaches for “gene prioritisation” [29,30] that use network information, which are now routinely used for filtering, for example, the large set of variants showing up in “genome-wide association analysis” (GWAS) studies. It also allows one to predict new genes potentially associated with the disease, which could, for example, be more “druggable” (e.g., “b” in Figure 1b). Another advantage is that locating the modules associated with a disease allows it to be related to one or more biological functions (e.g., biological/metabolic pathways, macromolecular complexes), due to the relationship between topological modules and functional modules described earlier. This gives further insight into the cellular basis of the disease. Finally, as the network topology and connections represent different molecular mechanisms, having a picture of the network context of the disease-related genes allows us to better understand those mechanisms (e.g., which gene activates which other one in cascades of transcription regulation) and eventually design therapeutic interventions aimed at, for example, rewiring a malfunctioning network (e.g., [31]).

Network approaches to human pathologies are not restricted to those molecular networks representing relationships between molecular entities. Many other disease-related data have been modelled as networks and studied from that point of view. These include drug–target relationships, drug–drug relationships, drug–disease associations, drug–side effect and disease–disease associations [32].

3. Phenotypes and Molecular Networks

To perform computational studies in general and systemic studies in particular on large collections of biomedical data, they have to be represented in a computer-tractable way, using standardised vocabularies (e.g., formal and standardised identifiers for genes, proteins) and ontologies (formal ways of representing knowledge, such as relationships between entities). In the case of phenotype data, there are multiple resources that provide these vocabularies and ontologies to describe different aspects of the human phenotypic landscape. A widely used vocabulary is that generated by the “Human Phenotype Ontology” (HPO) [33]. The HPO consists of a set of keywords describing human phenotypes related in a hierarchy, in which one can navigate from very general terms (e.g., “abnormality of the nervous system”) down to more specific ones (e.g., “seizure” → “non-motor seizure”).

Human diseases, characterised by a profile of distinctive HPO terms, can be clustered according to their phenotypic similarities. It has been known for a long time that many diseases and syndromes with similar phenotypes are caused by functionally related genes [21,34]. The extreme case is that of genetically heterogeneous diseases caused by genes involved in the same biological unit (such as a macromolecular complex, a pathway or process, or even an organelle). For example, the different types of Ehlers–Danlos syndromes are often caused by mutations in collagen genes or collagen interacting proteins that modify the structure, production or processing of collagen fibres. A large-scale study showed that genes causing the same (genetically heterogeneous) disease were frequently found to interact [35].

Beyond genetically heterogeneous diseases, several computational studies have observed that phenotypic similarity of otherwise distinct diseases is widely associated with shared protein interactions, both direct [36] as well as second-order interactions [37]. As mentioned above, diseases are organised in non-random modules in the human interactome, meaning that genes related to the same disease tend to be close in the interactome or even form a topological module [23]. The same modular organization observed for diseases has been quantified at the phenotype level [38], showing that genes associated with the same phenotype in genetic diseases tend to be close in the network of protein–protein interactions. Similarly, molecular-related phenotypes tend to overlap in the network, while unrelated phenotypes tend to be further apart.

4. Disease–Gene Predictions Using Phenotypic Descriptions and Molecular Networks

Early disease–gene prediction methods (e.g., [39,40]) were based on phenotypic descriptions of diseases, but did not use molecular networks. Direct protein interactions in human and model organisms were first used to predict novel genes associated with genetically heterogeneous diseases [41]. Lage et al. constructed a phenome–interactome network for human diseases, and used it to prioritise candidate genes [42]. In the absence at that time of curated phenotypic annotations such as those provided by the HPO, phenotypes were automatically extracted from OMIM records [43] using UMLS concepts [44]. UMLS defines a standardised vocabulary of medical concepts, and OMIM is a database of human genetic disorders. Since then, various studies have proposed a range of network-based methods to predict and prioritise novel disease-related genes [45], and several tools have been made available for this task (for a review see [30]).

Several approaches analyse a single molecular network (i.e., “homogeneous” network, with a single type of nodes and linkages). In these molecular network-based approaches, new genes are found in the proximity of known disease genes (“seed” genes). Algorithms that rely on global network distances, like random walk with restart (RWR) and network

diffusion, have been shown to outperform those that use local distances [46]. Integrated functional networks provide better results than single-type interaction networks [47].

In the case of complex, oligogenic or genetically heterogeneous diseases, the seed is usually the set of already known disease genes [48]. However, how can we define seed genes for patients with rare conditions that have no clear clinical diagnosis? A possible strategy is to compile seed genes from conditions with clinical manifestations (phenotypes) similar to those observed in the patient [27,49,50]. Using this strategy, methods that analyse the output of genome-wide sequencing platforms, such as Exomiser [51] and Phen-Gen [52], allow researchers to score candidate genes using a network-based approach. In these approaches, the clinical manifestations of patients should be provided using a controlled vocabulary, such as the HPO. These manifestations are then automatically compared to those of known diseases to construct a set of seed genes to start the network analysis. Several disease similarity metrics can be used to find phenotypically similar diseases (for a recent review see [53]).

Instead of using a single molecular network and establishing initial probabilities for seed genes based on phenotypic similarities, alternative approaches analyse multipartite (heterogeneous) networks. RWRH (random walk with restart on heterogeneous network) combines a disease–disease network (constructed based on phenotypic similarity) with the molecular network [54]. These two networks are linked by the connections between diseases and molecular nodes, generating a bipartite network. This approach was shown to outperform the equivalent method using a single molecular network in the task of assigning genes to diseases.

Disease–phenotype associations can be explicitly modelled as a bipartite network. This disease–phenotype network, together with parenthood relationships between phenotypes, a molecular network, and the corresponding inter-network links, establishes a tripartite network (diseases, phenotypes, and genes). This tripartite network, or context-sensitive network (CSN) has been shown to outperform RWRH in assigning genes to their corresponding diseases [55].

5. Phenotype–Gene Predictions

Similar methods to those described above aimed at predicting disease-related genes using molecular networks can be used to associate genes with phenotypes. The inverse problem, the prediction of phenotypes for a given gene/protein is outside the scope of this review (see CAFA challenge [56]).

Li et al. integrated protein interactions as well as disease and phenotype information to predict novel genes associated with phenotypes [57]. Their motivation was the study of the molecular mechanisms underlying phenotypes, which are the basis for personalised diagnosis and treatment in traditional Chinese medicine. Kahanda et al. used protein–protein interactions combined with other types of data, such as functional annotation and literature to predict phenotype–gene associations [58]. Another approach for assigning genes to phenotypes is that developed by Petegrosso et al., who used a transfer learning approach (tlDLP) on a tripartite network that contains HPO terms (and their parenthood relationships), Gene Ontology (GO) terms representing functional annotations (together with their ontological relationships) and a molecular network (protein–protein interactions) [59]. Gonzalez-Perez et al. [60] compared an RWR approach to a connectivity significance approach [19] using a single molecular network, obtaining overall better results with RWR. Yang et al. used a multipartite network embedding algorithm (LSGER) [61]. They obtained an overall improvement compared to FSGER (Fisher-based statistical model, as a baseline method) and PRINCE [27].

It is worth mentioning that the accuracy of phenotype–gene predictions depends not only on the method used, but also on several biomedical factors. This was shown in a recent study [60] in which the performance of RWR to predict genes associated with phenotypes was assessed and related to aspects like disease onset and pace of progression, phenotype prevalence, and gene product function. The biomedical factors that most

affected prediction performance were the monogenic or oligogenic nature of the disease, the type of phenotype, and the mode of inheritance. Better predictions were obtained for genes involved in oligogenic diseases (in contrast to monogenic diseases). Neoplasm phenotypes and abnormalities of the blood were among the best predicted phenotypes, in contrast to abnormalities of the eye or the nervous system that were poorly predicted. Phenotype–genes of autosomal-dominant inherited diseases obtained better global results than those with autosomal recessive inheritance.

There are other ways to associate genes with phenotypes that involve networks, although not necessarily molecular networks. For example, PhenFun [62] connects phenotypes with genes based on a cohort of patients with genomic disorders. This is achieved by identifying the genomic regions shared by multiple patients and connecting the genes that map to these regions to the phenotypes exhibited by these patients. Then, those phenotypes that are connected to the same gene via multiple patients are considered potentially associated. The strength and significance of the association is assessed using statistical methodology, as illustrated in Figure 2a and described in detail in [62]. They also used the genes associated with each phenotype to look for functional enrichment in terms of biological pathways. This was based on the assumption that it might be a specific pathway that leads to the phenotype, and alterations in different genes from the same pathway might lead to the same effects.

6. Phenotypes and Patient Stratification

One of the overarching themes of clinical research in the last decade has been the recognition that a one-size-fits-all disease label is often insufficient. It has been known for a long time that common diseases such as asthma are really composed of many subgroups or endotypes, characterised by differences in term of manifested phenotypes as well as often distinct underlying mechanisms [63]. A powerful way to discover subgroups within a disease is by stratifying patients. This can be done based purely on molecular data, such as using genetics and “omics” datasets to build gene signatures and model the underlying processes [64,65].

However, much can be achieved using phenotypic data. Patients should be well phenotyped, both in terms of breadth (ensuring all aspects of the patient’s disease is covered) and also depth (ensuring that the description of the disease is suitably precise). Most of the methods used to cluster patients are unsupervised approaches such as hierarchical clustering or related methods such as Partitioning Around Medoids (PAM), applied to a matrix of patients and factors.

It should be made clear that terminology is important here—there are studies that classify patients into different groups which they refer to as “phenotypes”, classifying based on factors such as levels of serum proteins [66] and scores on clinical tests [67]. In these studies, they are often talking about a small number of phenotypes. Conversely, in other work, phenotypes (such as HPO terms assigned to patients) are considered the factors themselves, and are used to perform clustering of the patients into groups.

In recent work, researchers looked in detail at how to assess phenotyping breadth and depth in a patient cohort [68]. Patients from three cohorts with very different properties were clustered based on their phenotypic profiles. For the cohorts with more inconsistent and less informative phenotypic data, it was found that many patients ended up in uninformative clusters, which included only a small number of very broad and un-specific phenotypes such as “intellectual disability”. Clearly, such broad groups are not sufficient or informative for stratifying patients in a useful manner, and it was suggested to remove uninformative patients from the dataset, given that filtering patients based on very minimal criteria (included patients must have more than two pathological phenotypes assigned) greatly changes the properties of the dataset and the grouping. Of course, if possible, further clinical examination can also improve the information available in patient phenotypic profiles [68].

Various methods exist for this purpose. Work in the field of pain has long used quantitative sensory testing (QST) for patient testing [69]; however, its utility is often questioned [70]. In recent work from Vollert et al., QST sensory profiles were used to stratify patients into one of three groups, “sensory loss” “thermal hyperalgesia” or “mechanical hyperalgesia” [67], using a deterministic algorithm based on features in their QST profile. Moreover, they also suggested a probabilistic algorithm that allowed a patient to potentially be assigned to more than one group.

Combining phenotype data with genetic data allows us to better understand the reasons for the stratification of patients. In seminal work from the BRIDGE-BPD consortium, looking at heritable bleeding and platelet disorders (BPD), Westbury et al. performed patient clustering on cohort members based on HPO data [71]. They found that the patient groups formed by the clustering corresponded to suspected BPD syndromes and pedigree membership. Patients with variants in specific genes also tended to cluster together, showing that patient stratification cannot only tell us how patients are related, but perhaps why, i.e., the underlying molecular/genetic mechanisms.

There has been increased interest in this area over recent years due to the COVID-19 pandemic. In a new and exciting study Mueller et al. used immune phenotypes for patient stratification [66]. More specifically, they measured serum cytokine levels for each patient and then performed clustering on the data, identifying three major groups of patients. Interestingly, group membership was highly correlated with the results of certain clinical tests.

Bringing together personalised medicine and molecular networks, there have also been several papers that perform network analysis at the individual level. Work by Liu et al. used expression data and protein interactions from the STRING database [72] to build sample-specific networks [73]. Other studies have focused on transcriptomic data, using linear interpolation in an attempt to reverse engineer single sample networks [74]. These personalised network approaches are set to improve greatly, as we move from bulk RNA sequencing to single-cell RNA sequencing, allowing a huge amount of information to be obtained at the individual level. Already, methods are being developed to analyse such data and produce networks down to the cellular level [75].

7. Phenotypes and Co-Morbidity

As well as stratifying patients, phenotype data from patient cohorts and disease databases can be used to identify patterns of co-dependency and comorbidity between phenotypes. There are methods to achieve this-based directly on phenotype co-occurrence within patient cohorts, or across many thousands or even millions of patients based on hospital records.

Many methods based on patient data use International Statistical Classification of Diseases and Related Health Problems (ICD) codes [76]. For example, in pioneering work, Hidalgo et al. used ICD-9 codes from thirteen million patients to identify comorbid disease relationships [77]. For each pair of diseases, they calculated relative risk of both being affected in the same patient to quantify comorbidity (Pearson’s correlation coefficient was also used). The pair information was then used to construct a Phenotypic Disease Network (PDN), which showed phenotypes belonging to specific disease families tending to be close in the PDN. This work has been heavily cited, and similar approaches have been applied to a range of other diseases, including ischemic heart disease [78], in which a PDN of comorbid phenotypes was built for sufferers of this disease. It has also been recently applied to look at comorbidity in depression [79], for which the authors used Pearson’s correlation to measure comorbidity between a large spectrum of phenotypic conditions. Both works used ICD-10 codes. Comorbidity networks can also be built between ICD-10 derived diseases. Such a method has been applied to type-2 diabetes, for example [80]. Other work has leveraged the relative risk calculation alongside temporal data to build disease progression networks, with clear potential use for diagnosis and prognosis [81].

ICD codes are widely used for creating hospital bills and insurance claims and as such are highly used and readily available. While ICD codes have the advantage of allowing the analysis of very large numbers of patients, the system has been criticised, largely due to potential bias related to their use in billing and other related administrative tasks [82]. The “human phenotype ontology” (HPO) described earlier provides an alternative system for assigning phenotypes, which due to its tree-like structure and other attributes, is well suited to comorbidity analysis. To this end, the PhenCo methodology was developed [83], which takes patient cohort data, consisting of phenotypic profiles for patients, annotated using the HPO. These profiles are then deconstructed to produce a bipartite network linking patients and profiles, such that it is possible to identify pairs of phenotypes that are shared by many patients (Figure 2b). Moreover, these pairs of phenotypes can be scored, such that those pairs that tend to occur together more often are scored highly.

These phenotype pairs by themselves are of interest as they help better understand phenotypic patterns within the cohort and formally quantify clinicians’ intuitions of how certain phenotypes tend to manifest together within their cohort. However, the real power of the method occurs when the pairs are combined to produce a phenotype network. PhenCo not only builds such a network, but implements an edge-based clustering method to obtain clusters of phenotypes that tend to occur together. Furthermore, it can integrate the phenotypic data with genomic data, such that phenotypes can be associated with functional systems, in a similar manner to the previously described PhenFun system [62]. This allows identification of clusters of phenotypes that overlap together, as well as clusters where the component phenotypes map to the same functional systems, providing insight into the molecular cause of the clustered phenotypes (Figure 2a).

There are also approaches that use HPO terms alongside data obtained directly from disease databases such as OMIM [43] and Orphanet [84]. For example, a recent study using the PhenoClusters methodology [85] gave insights into neuromuscular disease (NMD). The method obtains NMD related diseases from OMIM using keyword searching and expert manual curation. It then extracts NMD-related phenotypes by looking for HPO terms that are overrepresented within the NMD-related diseases. These are used to build a bipartite network, connecting phenotypes with diseases, and in a manner analogous to the PhenCo method described above, to find phenotypes that tend to occur together across diseases. These phenotype pairs are scored and can be assembled together to build networks that can in turn be clustered. Moreover, known disease–gene association data can be used to find clusters of related phenotypes that are associated with genes involved in similar pathways. Through collaboration with experts in the field of NMDs, it was possible to identify important clusters that can aid in differential diagnosis [85].

There are additional methods that, while not measuring comorbidity directly, attempt to look for relationships between phenotypes by incorporating additional data sources. In a study described earlier [38], similarities between phenotypes were quantified using the properties of the interactome. Similar work to this study combined HPO phenotypes with protein–protein interaction data, by mapping phenotypes to the interactome, and calculating similarity between the modules formed within the interactome by the distinct phenotypes [86].

As well as HPO terms and ICD codes, exciting approaches have also used symptom terms in the Medical Subject Headings (MeSH) [87] metadata fields of PubMed entries to build a human symptoms–disease network [37]. This idea has been extended to produce a tool that is capable of retrieving from the scientific literature relationships between HPO terms and terms from other annotation systems, including Gene Ontology as well as different biomedical ontologies [88]. However, this method differs in terms of the underlying approach, in that it searches for the terms within the abstracts of PubMed articles, rather than relying on the MeSH fields.

8. Resources

Most of the resources mentioned (e.g., databases and software tools) are free for academics and can be accessed on-line. In this section, we include a table with a summary of the main resources commented in this review as well as others related to the subject, including their web addressed (Table 1).

Table 1. Main online resources related to network approaches to diseases and phenotypes.

Name	Description	URL ¹	Reference
CytoScape	Widely used software for interactively representing and studying biological networks. Freely available for different operative systems	https://cytoscape.org/	[89]
STRING	Resource with networks of interactions and functional relationships between proteins in different organisms, inferred from different evidences	https://string-db.org/	[72]
Human Phenotype Ontology (HPO)	Controlled structured vocabulary for describing different aspects of human disease phenotypes/clinical signs	https://hpo.jax.org/app/	[33]
Online Mendelian Inheritance in Man (OMIM)	Catalogue of human genetic disorders and their related genes	https://www.omim.org/	[43]
Orphanet	Resource with information on rare diseases and orphan drugs	https://www.orpha.net/	[84]
Medical Subject Headings (MeSH)	Controlled vocabulary used to annotate PubMed bibliographic entries	https://www.ncbi.nlm.nih.gov/mesh/	[87]
CoMent	Relationships between biomedical concepts extracted from the literature	https://sysbiol.cnb.csic.es/CoMent/	[88]

¹ All URLs accessed on January 2022.

9. Discussion

Molecular biology revolutionised medicine by providing insight into the molecular mechanisms underlying pathological processes. The diagnosis and the treatment of many human diseases changed as molecular entities (e.g., genes, proteins, metabolites) started to be used as markers of a disease or targets to cure it. This molecular approach is highly “reductionist”, in the sense that a single, or a very small number of molecular entities, is assumed to be the cause (and target for an eventual treatment) of a disease. That assumption, fundamentally correct for some pathologies (e.g., monogenic diseases), has problems when applied to many complex pathologies such as cancer or Alzheimer’s. To complement these reductionist approaches, systems- and network-biology methods are being applied to the study of human pathologies, in what sometimes is called “Network Medicine”. The change of paradigm in these new approaches is that, due to their complexity, most diseases are better reflected at the level of complex networks, instead of single genes. Even in “classic” monogenic diseases, the causative gene is immersed in complex networks and hence, even if the onset of the disease depends on that single gene, other important factors, such as its severity or patient-specific manifestations depend on many other genes/mutations, requiring a more systemic approach for understanding them (e.g., cystic fibrosis [90]). From this point of view, diseases are seen as emergent properties of complex networks, which are

affected by both genetic and environmental factors, and as perturbations in the network structure (e.g., re-wiring) more than in the nodes (genes) themselves [8].

Approaching pathologies from a phenotypic point of view has some advantages with respect to the traditional disease-based point of view. Clinical signs can be closer to the molecular mechanism(s) behind the pathology, as they are more directly related to the phenotypic manifestations. Moreover, by definition, clinical signs are directly observable, and in many cases even quantifiable, and hence they can be reported for any pathological condition. Sometimes a particular disease cannot be diagnosed and only a collection of symptoms is reported. Clinical signs are also key for personalised medicine, as the patient symptom profiles can help in their stratification and the design of personalised interventions. For these reasons, network and systemic approaches to diseases are also gradually adopting this point of view and incorporating phenotypic information.

To incorporate phenotypic information, controlled vocabularies and ontologies for formally representing that kind of data are required, such as that developed by the HPO and other resources such as IDC or MeSH. Similarly, new methodologies for massively extracting phenotypic information from un-structured resources and representing them using those ontologies are also important (e.g., CoMent [88]). This would allow us to identify pathological phenotypes that tend to occur together across potentially millions of patients, in a similar manner to the previously mentioned methods PhenCo and PhenoClusters. Similarly, tools such as CoMent could potentially be extended to obtain co-occurrent phenotypes, as such pairs of phenotypes that tend to be mentioned together or not at all may represent potentially comorbid phenotypes.

To conclude, it is clear that the ongoing revolution in the generation of biological data, improvements in computational techniques for data analysis, particularly network analysis, and the ever-increasing recognition of the importance of deep phenotyping of patients [91], are providing us with a wealth of potential for better understanding disease and eventually finding better treatments.

Author Contributions: Writing—original draft preparation, writing—review and editing, J.A.G.R., J.P., M.C., E.D.-S. and F.P. All authors have read and agreed to the published version of the manuscript.

Funding: This work was partially funded by The Spanish Ministry of Economy and Competitiveness with European Regional Development Fund [grant numbers PID2019-108096RB-C21 and PID2019-108096RB-C22]; the European Food Safety Authority [grant number GP/EFSA/ENCO/2020/02]; the Andalusian Government with European Regional Development Fund [grant numbers UMA18-FEDERJA-102 and PAIDI 2020:PY20-00372]; Fundacion Progreso y Salud [grant number PI-0075-2017], also from the Andalusian Government; the Ramón Areces foundation, which funds project for the investigation of rare disease (National call for research on life and material sciences, XIX edition); the University of Malaga (Ayudas del I Plan Propio) and the Institute of Health Carlos III which funds the IMPaCT-Data project. The CIBERER is an initiative from the Institute of Health Carlos III. The conclusions, findings and opinions expressed in this scientific paper reflect only the view of the authors and not the official position of the European Food Safety Authority.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Mazzocchi, F. Complexity in Biology. Exceeding the Limits of Reductionism and Determinism Using Complexity Theory. *EMBO Rep.* **2008**, *9*, 10–14. [[CrossRef](#)]
2. Oltvai, Z.N.; Barabási, A.-L. Life's Complexity Pyramid. *Science* **2002**, *298*, 763–764. [[CrossRef](#)]
3. van Regenmortel, M.H. Reductionism and Complexity in Molecular Biology. Scientists Now Have the Tools to Unravel Biological and Overcome the Limitations of Reductionism. *EMBO Rep.* **2004**, *5*, 1016–1020. [[CrossRef](#)]
4. Avi Ma'ayan Complex Systems Biology. *J. R. Soc. Interface* **2017**, *14*, 20170391. [[CrossRef](#)]
5. Nurse, P. Systems Biology: Understanding Cells. *Nature* **2003**, *424*, 883. [[CrossRef](#)]
6. Kitano, H. Systems Biology: A Brief Overview. *Science* **2002**, *295*, 1662–1664. [[CrossRef](#)]

7. Barabasi, A.L.; Oltvai, Z.N. Network Biology: Understanding the Cell's Functional Organization. *Nat. Rev. Genet.* **2004**, *5*, 101–113. [[CrossRef](#)]
8. McGillivray, P.; Clarke, D.; Meyerson, W.; Zhang, J.; Lee, D.; Gu, M.; Kumar, S.; Zhou, H.; Gerstein, M.B. Network Analysis as a Grand Unifier in Biomedical Data Science. *Annu. Rev. Biomed. Data Sci.* **2018**, *1*, 153–180. [[CrossRef](#)]
9. Barabási, A.-L.; Gulbahce, N.; Loscalzo, J. Network Medicine: A Network-Based Approach to Human Disease. *Nat. Rev. Genet.* **2011**, *12*, 56–68. [[CrossRef](#)]
10. Furlong, L.I. Human Diseases through the Lens of Network Biology. *Trends Genet.* **2013**, *29*, 150–159. [[CrossRef](#)]
11. Chagoyen, M.; Ranea, J.A.; Pazos, F. Applications of Molecular Networks in Biomedicine. *Biol. Methods Protoc.* **2019**, *4*, bpz012. [[CrossRef](#)]
12. Zanzoni, A.; Soler-Lopez, M.; Aloy, P. A Network Medicine Approach to Human Disease. *FEBS Lett.* **2009**, *583*, 1759–1765. [[CrossRef](#)]
13. Mitra, K.; Carvunis, A.-R.; Ramesh, S.K.; Ideker, T. Integrative Approaches for Finding Modular Structure in Biological Networks. *Nat. Rev. Genet.* **2013**, *14*, 719–732. [[CrossRef](#)]
14. Gavin, A.C.; Aloy, P.; Grandi, P.; Krause, R.; Boesche, M.; Marzioch, M.; Rau, C.; Jensen, L.J.; Bastuck, S.; Dumpelfeld, B.; et al. Proteome Survey Reveals Modularity of the Yeast Cell Machinery. *Nature* **2006**, *440*, 631–636. [[CrossRef](#)]
15. Lin, C.-Y.; Lee, T.-L.; Chiu, Y.-Y.; Lin, Y.-W.; Lo, Y.-S.; Lin, C.-T.; Yang, J.-M. Module Organization and Variance in Protein-Protein Interaction Networks. *Sci. Rep.* **2015**, *5*, 9386. [[CrossRef](#)]
16. Hsia, C.-W.; Ho, M.-Y.; Shui, H.-A.; Tsai, C.-B.; Tseng, M.-J. Analysis of Dermal Papilla Cell Interactome Using STRING Database to Profile the Ex Vivo Hair Growth Inhibition Effect of a Vinca Alkaloid Drug, Colchicine. *Int. J. Mol. Sci.* **2015**, *16*, 3579–3598. [[CrossRef](#)]
17. Ravasz, E.; Somera, L.; Mongru, D.A.; Oltvai, Z.N.; Barabási, A.L. Hierarchical Organization of Modularity in Metabolic Networks. *Science* **2002**, *297*, 1551–1555. [[CrossRef](#)]
18. Freyre-González, J.A.; Alonso-Pavón, J.A.; Treviño-Quintanilla, L.G.; Collado-Vides, J. Functional Architecture of Escherichia Coli: New Insights Provided by a Natural Decomposition Approach. *Genome Biol.* **2008**, *9*, R154. [[CrossRef](#)]
19. Ghiassian, S.D.; Menche, J.; Barabási, A.-L. A Disease Module Detection (DIAMOND) Algorithm Derived from a Systematic Analysis of Connectivity Patterns of Disease Proteins in the Human Interactome. *PLoS Comput. Biol.* **2015**, *11*, e1004120. [[CrossRef](#)]
20. Shin, D.; Lee, J.; Gong, J.-R.; Cho, K.-H. Percolation Transition of Cooperative Mutational Effects in Colorectal Tumorigenesis. *Nat. Commun.* **2017**, *8*, 1270. [[CrossRef](#)]
21. Oti, M.; Brunner, H.G. The Modular Nature of Genetic Diseases. *Clin. Genet.* **2007**, *71*, 1–11. [[CrossRef](#)]
22. Rossin, E.J.; Lage, K.; Raychaudhuri, S.; Xavier, R.J.; Tatar, D.; Benita, Y.; Cotsapas, C.; Daly, M.J. Proteins Encoded in Genomic Regions Associated with Immune-Mediated Disease Physically Interact and Suggest Underlying Biology. *PLoS Genet.* **2011**, *7*, e1001273. [[CrossRef](#)]
23. Menche, J.; Sharma, A.; Kitsak, M.; Ghiassian, S.D.; Vidal, M.; Loscalzo, J.; Barabási, A.-L. Uncovering Disease-Disease Relationships through the Incomplete Interactome. *Science* **2015**, *347*, 1257601. [[CrossRef](#)]
24. Krishnan, A.; Zhang, R.; Yao, V.; Theesfeld, C.L.; Wong, A.K.; Tadych, A.; Volfovsky, N.; Packer, A.; Lash, A.; Troyanskaya, O.G. Genome-Wide Prediction and Functional Characterization of the Genetic Basis of Autism Spectrum Disorder. *Nat. Neurosci.* **2016**, *19*, 1454. Available online: <https://www.nature.com/articles/nn.4353#supplementary-information> (accessed on 1 January 2022). [[CrossRef](#)]
25. Cho, D.Y.; Kim, Y.A.; Przytycka, T.M. Chapter 5: Network Biology Approach to Complex Diseases. *PLoS Comp. Biol.* **2012**, *8*, e1002820. [[CrossRef](#)]
26. Cowen, L.; Ideker, T.; Raphael, B.J.; Sharan, R. Network Propagation: A Universal Amplifier of Genetic Associations. *Nature Rev. Genet.* **2017**, *18*, 551. [[CrossRef](#)]
27. Vanunu, O.; Magger, O.; Ruppín, E.; Shlomi, T.; Sharan, R. Associating Genes and Protein Complexes with Disease via Network Propagation. *PLoS Comput. Biol.* **2010**, *6*, e1000641. [[CrossRef](#)]
28. Kim, Y.A.; Przytycka, T.M. Bridging the Gap between Genotype and Phenotype via Network Approaches. *Front. Genet.* **2012**, *3*, 227. [[CrossRef](#)]
29. Moreau, Y.; Tranchevent, L.C. Computational Tools for Prioritizing Candidate Genes: Boosting Disease Gene Discovery. *Nat. Rev. Genet.* **2012**, *13*, 523–536. [[CrossRef](#)]
30. Zolotareva, O.; Kleine, M. A Survey of Gene Prioritization Tools for Mendelian and Complex Human Diseases. *J. Integr. Bioinform.* **2019**, *16*, 20180069. [[CrossRef](#)]
31. Lee, M.J.; Ye, A.S.; Gardino, A.K.; Heijink, A.M.; Sorger, P.K.; MacBeath, G.; Yaffe, M.B. Sequential Application of Anticancer Drugs Enhances Cell Death by Rewiring Apoptotic Signaling Networks. *Cell* **2012**, *149*, 780–794. [[CrossRef](#)] [[PubMed](#)]
32. Shahreza, M.L.; Ghadiri, N.; Mousavi, S.R.; Varshosaz, J.; Green, J.R. A Review of Network-Based Approaches to Drug Repositioning. *Brief. Bioinform.* **2017**, *19*, 878–892. [[CrossRef](#)] [[PubMed](#)]
33. Köhler, S.; Carmody, L.; Vasilevsky, N.; Jacobsen, J.O.B.; Danis, D.; Gouridine, J.-P.; Gargano, M.; Harris, N.L.; Matentzoglou, N.; McMurry, J.A.; et al. Expansion of the Human Phenotype Ontology (HPO) Knowledge Base and Resources. *Nucleic Acids Res.* **2019**, *47*, D1018–D1027. [[CrossRef](#)] [[PubMed](#)]
34. Brunner, H.G.; van Driel, M.A. From Syndrome Families to Functional Genomics. *Nat. Rev. Genet.* **2004**, *5*, 545–551. [[CrossRef](#)] [[PubMed](#)]

35. Goh, K.-I.; Cusick, M.E.; Valle, D.; Childs, B.; Vidal, M.; Barabási, A.-L. The Human Disease Network. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 8685–8690. [[CrossRef](#)]
36. Van Driel, M.A.; Bruggeman, J.; Vriend, G.; Brunner, H.G.; Leunissen, J.A.M. A Text-Mining Analysis of the Human Phenome. *Eur. J. Hum. Genet.* **2006**, *14*, 535–542. [[CrossRef](#)]
37. Zhou, X.; Menche, J.; Barabási, A.-L.; Sharma, A. Human Symptoms–Disease Network. *Nat. Commun.* **2014**, *5*, 4212. [[CrossRef](#)]
38. Chagoyen, M.; Pazos, F. Characterization of Clinical Signs in the Human Interactome. *Bioinformatics* **2016**, *32*, 1761–1765. [[CrossRef](#)]
39. Perez-Iratxeta, C.; Bork, P.; Andrade, M.A. Association of Genes to Genetically Inherited Diseases Using Data Mining. *Nat. Genet.* **2002**, *31*, 316–319. [[CrossRef](#)]
40. Freudenberg, J.; Propping, P. A Similarity-Based Method for Genome-Wide Prediction of Disease-Relevant Human Genes. *Bioinformatics* **2002**, *18* (Suppl. S2), S110–S115. [[CrossRef](#)]
41. Oti, M.; Snel, B.; Huynen, M.A.; Brunner, H.G. Predicting Disease Genes Using Protein-Protein Interactions. *J. Med. Genet.* **2006**, *43*, 691–698. [[CrossRef](#)] [[PubMed](#)]
42. Lage, K.; Karlberg, E.O.; Størling, Z.M.; Olason, P.I.; Pedersen, A.G.; Rigina, O.; Hinsby, A.M.; Tümer, Z.; Pociot, F.; Tommerup, N.; et al. A Human Phenome-Interactome Network of Protein Complexes Implicated in Genetic Disorders. *Nat. Biotechnol.* **2007**, *25*, 309–316. [[CrossRef](#)] [[PubMed](#)]
43. Hamosh, A.; Scott, A.F.; Amberger, J.S.; Bocchini, C.A.; McKusick, V.A. Online Mendelian Inheritance in Man (OMIM), a Knowledgebase of Human Genes and Genetic Disorders. *Nucleic Acids Res.* **2005**, *33*, D514–D517. [[CrossRef](#)] [[PubMed](#)]
44. Bodenreider, O. The Unified Medical Language System (UMLS): Integrating Biomedical Terminology. *Nucleic Acids Res.* **2004**, *32*, D267–D270. [[CrossRef](#)] [[PubMed](#)]
45. Wang, X.; Gulbahce, N.; Yu, H. Network-Based Methods for Human Disease Gene Prediction. *Brief Funct. Genom.* **2011**, *10*, 280–293. [[CrossRef](#)]
46. Navlakha, S.; Kingsford, C. The Power of Protein Interaction Networks for Associating Genes with Diseases. *Bioinformatics* **2010**, *26*, 1057–1063. [[CrossRef](#)]
47. Huang, J.K.; Carlin, D.E.; Yu, M.K.; Zhang, W.; Kreisberg, J.F.; Tamayo, P.; Ideker, T. Systematic Evaluation of Molecular Networks for Discovery of Disease Genes. *Cell Syst.* **2018**, *6*, 484–495. [[CrossRef](#)]
48. Köhler, S.; Bauer, S.; Horn, D.; Robinson, P.N. Walking the Interactome for Prioritization of Candidate Disease Genes. *Am. J. Hum. Genet.* **2008**, *82*, 949–958. [[CrossRef](#)]
49. Wu, X.; Jiang, R.; Zhang, M.Q.; Li, S. Network-Based Global Inference of Human Disease Genes. *Mol. Syst. Biol.* **2008**, *4*, 189. [[CrossRef](#)]
50. Cáceres, J.J.; Paccanaro, A. Disease Gene Prediction for Molecularly Uncharacterized Diseases. *PLoS Comput. Biol.* **2019**, *15*, e1007078. [[CrossRef](#)]
51. Smedley, D.; Jacobsen, J.O.; Jager, M.; Kohler, S.; Holtgrewe, M.; Schubach, M.; Siragusa, E.; Zemojtel, T.; Buske, O.J.; Washington, N.L.; et al. Next-Generation Diagnostics and Disease-Genes Discovery with the Exomiser. *Nat. Protoc.* **2015**, *10*, 2004–2015. [[CrossRef](#)]
52. Javed, A.; Agrawal, S.; Ng, P.C. Phen-Gen: Combining Phenotype and Genotype to Analyze Rare Disorders. *Nat. Methods* **2014**, *11*, 935–937. [[CrossRef](#)]
53. Cheng, L.; Zhao, H.; Wang, P.; Zhou, W.; Luo, M.; Li, T.; Han, J.; Liu, S.; Jiang, Q. Computational Methods for Identifying Similar Diseases. *Mol. Ther. Nucleic Acids* **2019**, *18*, 590–604. [[CrossRef](#)]
54. Li, Y.; Patra, J.C. Genome-Wide Inferring Gene-Phenotype Relationship by Walking on the Heterogeneous Network. *Bioinformatics* **2010**, *26*, 1219–1224. [[CrossRef](#)]
55. Chen, Y.; Xu, R. Context-Sensitive Network-Based Disease Genetics Prediction and Its Implications in Drug Discovery. *Bioinformatics* **2017**, *33*, 1031–1039. [[CrossRef](#)]
56. Zhou, N.; Jiang, Y.; Bergquist, T.R.; Lee, A.J.; Kacsoh, B.Z.; Crocker, A.W.; Lewis, K.A.; Georghiou, G.; Nguyen, H.N.; Hamid, M.N.; et al. The CAFA Challenge Reports Improved Protein Function Prediction and New Functional Annotations for Hundreds of Genes through Experimental Screens. *Genome Biol.* **2019**, *20*, 244. [[CrossRef](#)]
57. Li, X.; Zhou, X.; Peng, Y.; Liu, B.; Zhang, R.; Hu, J.; Yu, J.; Jia, C.; Sun, C. Network Based Integrated Analysis of Phenotype-Genotype Data for Prioritization of Candidate Symptom Genes. *BioMed Res. Int.* **2014**, *2014*, 10. [[CrossRef](#)]
58. Kahanda, I.; Funk, C.; Verspoor, K.; Ben-Hur, A. PHENOstruct: Prediction of Human Phenotype Ontology Terms Using Heterogeneous Data Sources. *F1000Research* **2015**, *4*, 259. [[CrossRef](#)]
59. Petegrosso, R.; Park, S.; Hwang, T.H.; Kuang, R. Transfer Learning across Ontologies for Phenome-Genome Association Prediction. *Bioinformatics* **2017**, *33*, 529–536. [[CrossRef](#)]
60. Gonzalez-Perez, S.; Pazos, F.; Chagoyen, M. Factors Affecting Interactome-Based Prediction of Human Genes Associated with Clinical Signs. *BMC Bioinform.* **2017**, *18*, 340. [[CrossRef](#)]
61. Yang, K.; Wang, N.; Liu, G.; Wang, R.; Yu, J.; Zhang, R.; Chen, J.; Zhou, X. Heterogeneous Network Embedding for Identifying Symptom Candidate Genes. *J. Am. Med. Inform. Assoc.* **2018**, *25*, 1452–1459. [[CrossRef](#)] [[PubMed](#)]
62. Jabato, F.M.; Seoane, P.; Perkins, J.R.; Rojano, E.; García Moreno, A.; Chagoyen, M.; Pazos, F.; Ranea, J.A.G. Systematic Identification of Genetic Systems Associated with Phenotypes in Patients with Rare Genomic Copy Number Variations. *Hum. Genet.* **2021**, *140*, 457–475. [[CrossRef](#)] [[PubMed](#)]

63. Lötvall, J.; Akdis, C.A.; Bacharier, L.B.; Bjermer, L.; Casale, T.B.; Custovic, A.; Lemanske, R.F., Jr.; Wardlaw, A.J.; Wenzel, S.E.; Greenberger, P.A. Asthma Endotypes: A New Approach to Classification of Disease Entities within the Asthma Syndrome. *J. Allergy Clin. Immunol.* **2011**, *127*, 355–360. [[CrossRef](#)] [[PubMed](#)]
64. Chen, P.; Fan, Y.; Man, T.-K.; Hung, Y.S.; Lau, C.C.; Wong, S.T.C. A Gene Signature Based Method for Identifying Subtypes and Subtype-Specific Drivers in Cancer with an Application to Medulloblastoma. *BMC Bioinform.* **2013**, *14* (Suppl. S18), S1. [[CrossRef](#)] [[PubMed](#)]
65. Lundberg, A.; Lindström, L.S.; Harrell, J.C.; Falato, C.; Carlson, J.W.; Wright, P.K.; Foukakis, T.; Perou, C.M.; Czene, K.; Bergh, J.; et al. Gene Expression Signatures and Immunohistochemical Subtypes Add Prognostic Value to Each Other in Breast Cancer Cohorts. *Clin. Cancer Res.* **2017**, *23*, 7512–7520. [[CrossRef](#)]
66. Mueller, Y.M.; Schrama, T.J.; Ruijten, R.; Schreurs, M.W.J.; Grashof, D.G.B.; van de Werken, H.J.G.; Lasinio, G.J.; Álvarez-Sierra, D.; Kiernan, C.H.; Castro Eiro, M.D.; et al. Stratification of Hospitalized COVID-19 Patients into Clinical Severity Progression Groups by Immuno-Phenotyping and Machine Learning. *Nat. Commun.* **2022**, *13*, 915. [[CrossRef](#)]
67. Vollert, J.; Maier, C.; Attal, N.; Bennett, D.L.H.; Bouhassira, D.; Enax-Krumova, E.K.; Finnerup, N.B.; Freynhagen, R.; Gierthmühlen, J.; Haanpää, M.; et al. Stratifying Patients with Peripheral Neuropathic Pain Based on Sensory Profiles: Algorithm and Sample Size Recommendations. *Pain* **2017**, *158*, 1446–1455. [[CrossRef](#)]
68. Rojano, E.; Córdoba-Caballero, J.; Jabato, F.M.; Gallego, D.; Serrano, M.; Pérez, B.; Parés-Aguilar, Á.; Perkins, J.R.; Ranea, J.A.G.; Seoane-Zonjic, P. Evaluating, Filtering and Clustering Genetic Disease Cohorts Based on Human Phenotype Ontology Data with Cohort Analyzer. *J. Pers. Med.* **2021**, *11*, 730. [[CrossRef](#)]
69. Arendt-Nielsen, L.; Yarnitsky, D. Experimental and Clinical Applications of Quantitative Sensory Testing Applied to Skin, Muscles and Viscera. *J. Pain* **2009**, *10*, 556–572. [[CrossRef](#)]
70. Krumova, E.K.; Geber, C.; Westermann, A.; Maier, C. Neuropathic Pain: Is Quantitative Sensory Testing Helpful? *Curr. Diabetes Rep.* **2012**, *12*, 393–402. [[CrossRef](#)]
71. Westbury, S.K.; Turro, E.; Greene, D.; Lentaigne, C.; Kelly, A.M.; Bariana, T.K.; Simeoni, I.; Pillois, X.; Attwood, A.; Austin, S.; et al. Human Phenotype Ontology Annotation and Cluster Analysis to Unravel Genetic Defects in 707 Cases with Unexplained Bleeding and Platelet Disorders. *Genome Med.* **2015**, *7*, 36. [[CrossRef](#)] [[PubMed](#)]
72. Szklarczyk, D.; Gable, A.L.; Lyon, D.; Junge, A.; Wyder, S.; Huerta-Cepas, J.; Simonovic, M.; Doncheva, N.T.; Morris, J.H.; Bork, P.; et al. STRING V11: Protein-Protein Association Networks with Increased Coverage, Supporting Functional Discovery in Genome-Wide Experimental Datasets. *Nucleic Acids Res.* **2019**, *47*, D607–D613. [[CrossRef](#)] [[PubMed](#)]
73. Liu, X.; Wang, Y.; Ji, H.; Aihara, K.; Chen, L. Personalized Characterization of Diseases Using Sample-Specific Networks. *Nucleic Acids Res.* **2016**, *44*, e164. [[CrossRef](#)] [[PubMed](#)]
74. Kuijjer, M.L.; Tung, M.G.; Yuan, G.; Quackenbush, J.; Glass, K. Estimating Sample-Specific Regulatory Networks. *iScience* **2019**, *14*, 226–240. [[CrossRef](#)]
75. Dai, H.; Li, L.; Zeng, T.; Chen, L. Cell-Specific Network Constructed by Single-Cell RNA Sequencing Data. *Nucleic Acids Res.* **2019**, *47*, e62. [[CrossRef](#)]
76. Harrison, J.E.; Weber, S.; Jakob, R.; Chute, C.G. ICD-11: An International Classification of Diseases for the Twenty-First Century. *BMC Med. Inform. Decis. Mak.* **2021**, *21*, 206. [[CrossRef](#)]
77. Hidalgo, C.A.; Blumm, N.; Barabási, A.-L.; Christakis, N.A. A Dynamic Network Approach for the Study of Human Phenotypes. *PLoS Comput. Biol.* **2009**, *5*, e1000353. [[CrossRef](#)]
78. Zhou, D.; Wang, L.; Ding, S.; Shen, M.; Qiu, H. Phenotypic Disease Network Analysis to Identify Comorbidity Patterns in Hospitalized Patients with Ischemic Heart Disease Using Large-Scale Administrative Data. *Healthcare* **2022**, *10*, 80. [[CrossRef](#)]
79. Qiu, H.; Wang, L.; Zeng, X.; Pan, J. Comorbidity Patterns in Depression: A Disease Network Analysis Using Regional Hospital Discharge Records. *J. Affect. Disord.* **2022**, *296*, 418–427. [[CrossRef](#)]
80. Khan, A.; Uddin, S.; Srinivasan, U. Comorbidity Network for Chronic Disease: A Novel Approach to Understand Type 2 Diabetes Progression. *Int. J. Med. Inform.* **2018**, *115*, 1–9. [[CrossRef](#)]
81. Jeong, E.; Ko, K.; Oh, S.; Han, H.W. Network-Based Analysis of Diagnosis Progression Patterns Using Claims Data. *Sci. Rep.* **2017**, *7*, 15561. [[CrossRef](#)]
82. Liebovitz, D.M.; Fahrenbach, J. COUNTERPOINT: Is ICD-10 Diagnosis Coding Important in the Era of Big Data? No. *Chest* **2018**, *153*, 1095–1098. [[CrossRef](#)] [[PubMed](#)]
83. Díaz-Santiago, E.; Jabato, F.M.; Rojano, E.; Seoane, P.; Pazos, F.; Perkins, J.R.; Ranea, J.A.G. Phenotype-Genotype Comorbidity Analysis of Patients with Rare Disorders Provides Insight into Their Pathological and Molecular Bases. *PLoS Genet.* **2020**, *16*, e1009054. [[CrossRef](#)] [[PubMed](#)]
84. Pavan, S.; Rommel, K.; Mateo Marquina, M.E.; Höhn, S.; Lanneau, V.; Rath, A. Clinical Practice Guidelines for Rare Diseases: The Orphanet Database. *PLoS ONE* **2017**, *12*, e0170365. [[CrossRef](#)] [[PubMed](#)]
85. Díaz-Santiago, E.; Claros, M.G.; Yahyaoui, R.; de Diego-Otero, Y.; Calvo, R.; Hoenicka, J.; Palau, F.; Ranea, J.A.G.; Perkins, J.R. Decoding Neuromuscular Disorders Using Phenotypic Clusters Obtained from Co-Occurrence Networks. *Front. Mol. Biosci.* **2021**, *8*, 635074. [[CrossRef](#)] [[PubMed](#)]
86. Peng, J.; Hui, W.; Shang, X. Measuring Phenotype-Phenotype Similarity through the Interactome. *BMC Bioinform.* **2018**, *19*, 114. [[CrossRef](#)]

87. Lowe, H.J.; Barnett, G.O. Understanding and Using the Medical Subject Headings (MeSH) Vocabulary to Perform Literature Searches. *JAMA* **1994**, *271*, 1103–1108. [[CrossRef](#)]
88. Pazos, F.; Chagoyen, M.; Seoane, P.; Ranea, J.A.G. CoMent: Relationships Between Biomedical Concepts Inferred From the Scientific Literature. *J. Mol. Biol.* **2022**, *434*, 167568. [[CrossRef](#)]
89. Smoot, M.E.; Ono, K.; Ruscheinski, J.; Wang, P.-L.; Ideker, T. Cytoscape 2.8: New Features for Data Integration and Network Visualization. *Bioinformatics* **2011**, *27*, 431–432. [[CrossRef](#)]
90. Guggino, W.B.; Stanton, B.A. New Insights into Cystic Fibrosis: Molecular Switches That Regulate CFTR. *Nat. Rev. Mol. Cell Biol.* **2006**, *7*, 426. [[CrossRef](#)]
91. Wright, J.T.; Herzberg, M.C. Science for the next Century: Deep Phenotyping. *J. Dent. Res.* **2021**, *100*, 785–789. [[CrossRef](#)] [[PubMed](#)]