Biochemical and Biophysical Research Communications 445 (2014) 757–773

Contents lists available at ScienceDirect

Biochemical and Biophysical Research Communications

journal homepage: www.elsevier.com/locate/ybbrc

Review

Integration, visualization and analysis of human interactome

Chiara Pastrello [a,1], Elisa Pasini [b,1], Max Kotlyar [a], David Otasek [a], Serene Wong [a,c], Waheed Sangrar [a], Sara Rahmati [a,d], Igor Jurisica [a,d,e,*]

[a] Princess Margaret Cancer Centre, University Health Network and TECHNA Institute for the Advancement of Technology for Health, TMDT, Room 11–314, 101 College Street, Toronto, ON M5G 1L7, Canada
[b] CRO Aviano National Cancer Institute, Cancer Bio-Immunotherapy Unit, Via Franco Gallini 2, 33081 Aviano, PN, Italy
[c] York University, Department of Computer Science and Engineering, Toronto, ON, Canada
[d] University of Toronto, Department of Medical Biophysics, Toronto, ON, Canada
[e] University of Toronto, Department of Computer Science, Toronto, ON, Canada

ARTICLE INFO

Article history:
Available online 1 February 2014

Keywords:
Omics
Interactome
Network analysis
Gastric cancer
Visual data mining

ABSTRACT

Data integration and visualization are crucial to obtain meaningful hypotheses from the diversity of 'omics' fields and the large volume of heterogeneous and distributed data sets. In this review we focus on network analysis as a key technique to integrate, visualize and extrapolate relevant information from diverse data. We first describe challenges in integrating different types of data and then focus on systematically exploring network properties to gain insight into network function. We also describe the relationship between network structures and function of elements that form it. Next, we highlight the role of the interactome in connecting data derived from different experiments, and we stress the importance of network analysis to recognize interaction context-specific features. Finally, we present an example integration to demonstrate the value of the network approach in cancer research, and highlight the importance of dynamic data in the specific context of signaling pathways.

Contents

1. Omics . . . 758
2. Integration . . . 758
3. Accurate representation of omics data . . . 758
   3.1. Data collection and storage . . . 758
   3.2. Data exchange . . . 759
   3.3. Example difficulties in data exchange . . . 759
4. Network visualization . . . 759
5. Example integration . . . 760
   5.1. New era in pharmacology . . . 760
   5.2. Gastric cancer . . . 760
   5.3. Illustrative network for gastric cancer . . . 761
   5.4. Investigating the relationship between tetracycline and gastric cancer through network analysis . . . 761
6. Network structure–function relationship . . . 762
   6.1. Network properties . . . 762
      6.1.1. Global network properties . . . 763
      6.1.2. Local network properties . . . 763
   6.2. Computational challenges . . . 764
7. Importance of the interactome in network analysis . . . 764
   7.1. Expansion of the known human interactome . . . 764

* Corresponding author at: Princess Margaret Cancer Centre, University Health Network, TMDT, Room 11–314, 101 College Street, Toronto, ON M5G 1L7, Canada.
E-mail address: juris@ai.utoronto.ca (I. Jurisica).
[1] These authors contributed equally to this work.

http://dx.doi.org/10.1016/j.bbrc.2014.01.151
0006-291X/

7.2. Interactome expansion . . . 764
7.3. Benefits of interactome expansion . . . 764
7.4. Interactome biases . . . 764
7.5. Interactome accuracy . . . 765
7.6. Comparisons of human interactomes across time . . . 765
8. Power of network analysis to identify context-specific information . . . 767
8.1. Context-specificity of interactomes . . . 767
8.2. A context application: modeling of signaling pathways . . . 768
9. Discussion . . . 769
Acknowledgments . . . 769
References . . . 769

1. Omics

The suffix "omics" is appended to words describing a field of study and usually involves large scale, comprehensive and systematic techniques. The first use of omics in this manner was genomics [1], and the first international project to collect the most complete data set, a building block for genomics was the Human Genome Project, launched in 1990 and completed in 2003 (http://www.genome.gov/10001772). The list of completed genomes includes 170 Eukaryota organisms, almost 3500 Virus and more than 2500 Bacteria (as listed in http://www.ebi.ac.uk/genomes/; last accessed 13th Dec. 2013). Comparably, there are 277 completed proteomes for Eukaryota organisms, more than 1600 Bacteria and more than 1100 Virus (as listed in http://www.uniprot.org/taxonomy/complete-proteomes; last accessed 13th Dec. 2013). Moreover, the structure of more than 94,000 proteins across different organisms has been described so far (as listed in http://www.rcsb.org/pdb/home/home.do; last accessed 13th Dec. 2013; the organism with the highest number of structures being *Homo Sapiens*).

Increasing data collections enable more comprehensive analyses, but coping with this data deluge is not trivial. For example, one mass spectrometry experiment can result in thousands to hundreds of thousands of spectra for one sample [2]. Likewise, next-generation sequencing produces millions of reads per sample [3]. The amount of data being generated calls for uniformity, standardization and optimized workflows [4]. Even if these are very basic concepts, they are not as widespread as one would expect. For example, there are 338 protein–protein interaction (PPI) databases, 243 metabolic pathway databases and 202 signaling pathway databases (as listed in http://www.pathguide.org/; last accessed 13th Dec. 2013) of which only some are in a format that supports data interchange across databases. Navigating through these vast resources can be challenging but integrating such data is both beneficial and increasingly necessary.

2. Integration

The emergence of high-throughput (HT) assays shifted research from hypothesis-driven exploration to data-driven hypothesis generation. However, generating substantially more data, HT methods in turn led to shifting from predominantly using statistical tools to depending on computational biology approaches, especially data mining and machine learning algorithms, to aid data analysis and interpretation [5,6]. As the number of omics disciplines grows, and with them the amount of data, the combination of an increasing number of different perspectives can give the scientist a more complete (and more realistic) view of the system they are studying. Now, the challenge is data integration, and in turn integrative data analysis [7]. For example, integrating gene expression with copy number variation data, mutation status, methylation profile and microRNA targeting can highlight the key players in a specific disease. Complex, multifactorial diseases can only be fully investigated with this type of approach. Relationships between these data and entities can effectively be represented as graphs. Thus, network visualization and analysis is becoming one of the key tools for integrative analysis.

3. Accurate representation of omics data

Data integration requires immense attention to information representation, annotation and support for accurate data exchange. Integrative computational biology supports modeling biological processes using data integrated across many omics fields. To address this, numerous data architectures have been established to effectively and efficiently collect, store, annotate and exchange data. These architectures vary in scope, intent, and standards they use. They are continuously being updated to represent the most current knowledge, and as such will contain inconsistencies and incompletions. Often, researchers rely on one or many such architectures to integrate pre-existing research with their own, or to share their own results. Understanding the nature of the architectures available as well as being able to accurately specify which have been used is critical to reducing ambiguities in this process, and improves the quality and utility of published results.

3.1. Data collection and storage

Information in omics data changes frequently, arriving in the form of peer-reviewed studies. These can be small-scale, hypothesis-based studies with a small number of results, or wider scale HT studies with thousands of results. Collecting and storing these results has necessitated the use of many diverse omics databases. These databases vary greatly in scope; for our purposes, they address entities, relationships between them, and annotations. Entity databases cover proteins, genes, small molecules, or other biologically relevant objects, and include for example UniProt [8], GeneCards [9], RefSeq [10], and DrugBank [11,12]. Relationship databases describe how these entities relate to each other. These include, but are not limited to, protein or gene interactions, drug targeting, and biological pathways. Some examples include IntAct [13], I2D [14–16], BioGRID [17], and KEGG [18]. IMEx consortium [19] and PSICQUIC registry [20] are notable collections of interaction databases that share representation and curation workflow standards [19]. Annotation databases attempt to create indices of terms and definitions. These terms are intended to unambiguously describe entities or relationships between them, often structured as an ontology, which further describes how terms relate to each other in a standard way. Some examples include Gene Ontology [21] and the numerous controlled vocabularies included in the PSI-MI standard [22]. Some databases may rely on each other's standards: for example, UniProt and other sequence databases may contain references to relevant Gene Ontology terms in their records. Resources such as GeneCards [23] integrate heterogeneous

and distributed databases for individual human genes, providing a one-stop-shop for gene annotation.

The content of a specific database relates directly to the criteria by which results are accepted. This varies from database to database. Some may rely on human curation, which can have numerous schemes, from a single reviewer to a requirement for several independent reviewers to agree on data to include. Others may be entirely dependent on computational methods, which vary in accuracy. The consequence of these differing standards and methodologies is that there is inter-database disagreement. This is evident even in the conflicting sizes of omics databases; for example, UniProt's current release contains 20,274 *H. Sapiens* SwissProt entries, which are reviewed before inclusion, as well as 114,645 *H. Sapiens* TrEMBL entries, which are unreviewed and automatically annotated. In contrast RefSeq contains 68,897 *H. Sapiens* proteins, both reviewed and unreviewed.

Another consequence of differing standards is that stored data is heterogeneous between databases. For example, UniProt and RefSeq entries are organized differently, and contain different information. Accessing data for EGFR in UniProt and in RefSeq in FASTA format, a standard for sequence [24], shows that while the sequences match, different annotations have been included. Differences such as these can be identified across all omics databases. Researchers need to be aware that there is no 'gold standard', and pay close attention to how information is entered into the databases they may be using.

3.2. Data exchange

Being able to access and interact with data requires that it be provided in an agreed upon standard. Many standards exist in the omics disciplines, each shaped by the needs of researchers and curators.

Simple information exchange, such as sharing a list of molecules or molecule pairs, can be done through the use of database identifiers. Most databases make use of unique identifiers, such as the UniProt ID, Entrez ID or GO ID, to reference their records. These identifiers are a desirable way to reference entities, relationships, and annotations. However, before selecting what type of identifier to use, the data architecture in which the originating database operates must be understood; many records in a database may be superficially similar but referenced by different identifiers. In addition, updates to the record may happen over time, requiring the researcher to know which version of the database they were using to get the same data in the future. Examples of these difficulties will be examined in more detail below.

Lists of IDs are easy to edit and curate, but lack descriptive ability. As the need for accurate descriptions of the data being exchanged increases, there is a corresponding increase in the complexity of the data format being used. This increase in complexity makes it harder to edit and create data, and often necessitates the use of specialized software tools, such as validators and editors, to ease the process. Some such formats are PSI-MI, BioPAX [25], KEGG's KGML [26], and SBML [27]. All of these formats were developed to address specific needs, and these needs should be examined before determining whether or not that format is appropriate for a particular exchange. PSI-MI [28], developed by the HUPO Proteomics Standards Initiative, is an XML format that standardizes the exchange of molecular interaction data. In comparison, BioPAX and KEGG also contain molecular interactions, but differ in that they go beyond binary molecular interactions, including interactions that require more than two interactors, or may involve catalysts, or inhibitors. This is an important and significant difference.

Intelligently choosing which standard to use for data exchange is a process that requires familiarizing oneself with the options available and comparing these with the needs of that exchange.

3.3. Example difficulties in data exchange

Neglecting to pay attention to data architecture in the research process can lead to difficulties, particularly in the replicability of reported results. To illustrate, we will examine the many possible interpretations of a single protein – EGFR, epidermal growth factor receptor. The short name is convenient, but can introduce problems very quickly. For example, a search for EGFR in UniProt yields 409 genes and 42 proteins. There are multiple reasons why this occurs; proteins that exist across multiple species share the same name and variants of a protein do as well. The most accurate, unambiguous descriptor would be the protein's sequence, but this is cumbersome to use. Each UniProt record contains a sequence and has a unique UniProt ID. A UniProt ID, for example P00533, a human version of EGFR, is an efficient and almost unambiguous way to reference a protein. As our understanding of what sequence constitutes a protein also changes over time; however, such IDs become more ambiguous. When UniProt undergoes periodic updates, the information associated with individual proteins as well as the sequences that characterize them may change. P00533 had undergone 200 revisions since its first appearance in 1988, in version 9.0 of UniProt. Our reference to P00533 was retrieved from UniProt version 2013_11. Documenting the version of a resource used is an often overlooked but important step in data exchange.

Competing standards can make exchanges frustrating. For example, a researcher might get a list of molecules of interest from a colleague using NCBI Locus IDs and be asked to find relevant drugs that target those molecules from the DrugBank database. DrugBank records store targets using UniProt and GeneCard identifiers. In UniProt, the entry for our example protein P00533, contains 24 different relevant NCBI Locus IDs. Selecting from this set of IDs manually using expert knowledge is the most accurate way to translate between different database identifiers, but this is time intensive and impractical, particularly when dealing with hundreds or thousands of identifiers at a time. Some automatic tools, such as DAVID [29], have been built to automate these annotation and ID mapping processes. Results will vary from tool to tool, even in expert manual translation; similarly for proteomics databases, 'gold-standard' accuracy may not be possible.

These difficulties only cover a small portion of the confusion that is possible when dealing with even the simplest forms of data exchange. With complex data such as PSI-MI, these difficulties can increase dramatically. Understanding that such issues may exist with omics data, and becoming familiar with the nature of databases and standards, user can limit ambiguity or confusion, and help make results more easily replicable. Importantly, as documented in the Retraction Watch (http://retractionwatch.com/) many, even high impact papers get retracted due to 'sloppy data handling' or 'inability to replicate results'. It is alarming, since we are experiencing an exponential increase in retracted papers due to these issues (Fig. 1A). Moreover, as highlighted in Fig. 1B, patients seem to be the most affected by retracted publications.

4. Network visualization

Networks comprise nodes and edges, where individual entities (e.g., genes, proteins, drugs) are represented by nodes, and relationships between components are represented by edges (e.g., phosphorylation, drug-targeting, transcriptional regulation). Network visualization and analysis are powerful tools for integrating diverse data, and collecting and highlighting useful information, in order to obtain an insightful representation of the system being studied [30]. Network data integration is not immune to pitfalls, either at the data integration level or at the network analysis one. More than ever before, the variety and quantity of data continues to increase, and the tasks that need to be performed

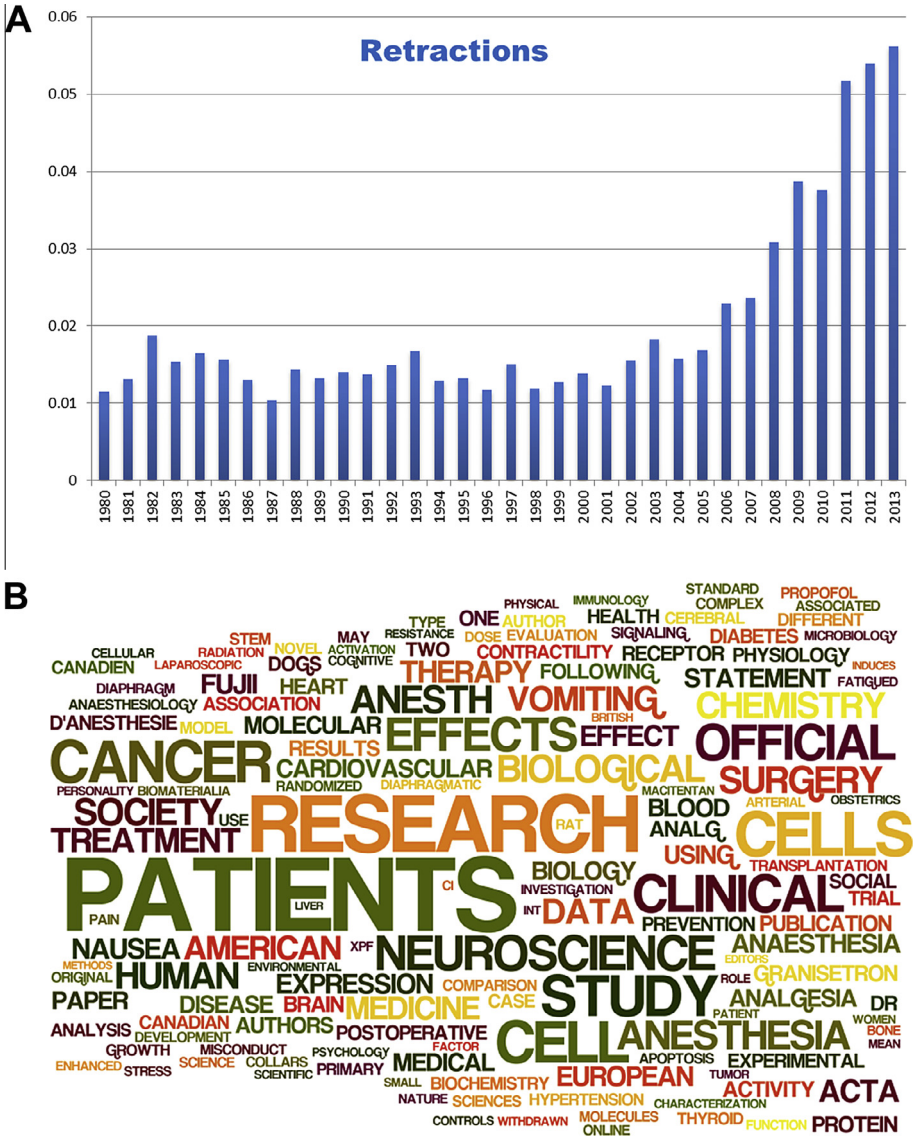

Fig. 1. Retracted papers from 1980 till 2013. (A) The figure shows the percentage of papers retracted over papers published for the same year. (B) The figure shows the frequency of words in titles of retracted journals (created with Wordle http://www.wordle.net/).

are and will be changing. This highlights the need for network visualization tools to be versatile and adaptive to change with data and researcher's needs [31].

Several tools have been developed to aid in managing big networks and extracting meaningful information from integrated data, some more focused on molecular tasks and others more flexible (recently compared in [32]). In this review we will show how to integrate and visualize different types of data using NAViGaTOR (http://ophid.utoronto.ca/navigator), a scalable, interactive network visualization and analysis system. To highlight the power of integrative network analysis we collected diverse datasets on *Helicobacter pylori*-derived gastric cancer, comprising gene expression, methylation, copy number variation, microRNAs, drugs, SNPs, protein interactions and pathways.

5. Example integration

5.1. New era in pharmacology

For several years the 'one-target, one-drug' paradigm has been the principle driving research and development in different areas of medicine and pharmaceutical sciences [33]. In recent years, new technologies such as HT gene expression profiling [34] have increased the molecular characterization of differences detectable between healthy and disease individuals, allowing the identification of a rising number of disease-related genes [35,36]. However, we can observe limitations of genomic approaches to design new therapies; in fact these methods usually deal with information about changes in the expression level of individual genes. With the recent advances in the omics area, viewing a drug target as part of a complex network of interacting proteins, and modeling the interaction of the drug with that entire network is a promising approach to manage the massive amounts of data and to improve our understanding of the impact of chemicals in human health. In this part of the review we present an example of a multifactorial disease, such as gastric cancer.

5.2. Gastric cancer

Gastric cancer (GC) is the second most common cause of cancer death worldwide [37], although the prevalence of gastric carcinoma varies in different parts of the world [38] and among various

ethnic groups [39]. The etiology of gastric cancer is multifactorial including both dietary (alcohol, nitrates and salt intake) [40,41] and non-dietary factors, including smoking, family history and occupation [42–44].

Epidemiological evidence shows that the bacterium, *H. pylori*, which infects 50% of the world's population, leads to the development of diffuse type gastric cancer in at least 1% of infected individuals [45,46]. The link between *H. pylori* infection and GC is mainly constituted by chronic inflammation, which is recognized as a major step in the initiation and development of several tumors, including GC [47]. Moreover the microorganism produces a variety of virulence factors (CagA, VacA) that may alter host intracellular signaling pathways and lead to neoplastic transformation [48,49]. The development of gastric cancer is a complex, multistep process starting from an inflammatory process with the subsequent accumulation of multiple genetic and epigenetic alterations in gastric epithelia involving oncogenes, tumor suppressor genes, DNA repair genes, cell cycle regulators, and signaling molecules. Besides bacterial-mediated effects on cell proliferation and DNA integrity, *H. pylori*-induced aberrant DNA methylation emerges as an important mechanism in stomach carcinogenesis [50,51]. Accumulating data have demonstrated that *H. pylori* could also induce the production of microRNAs [52], known to be crucial epigenetic regulators of immune responses [53] and cancer development [54].

The identification of host genetic components associated with the risk of this disease underlines that host factors can also influence the pathogenic activity of the bacterium. The presence of single nucleotide polymorphisms (SNPs) in several genes, including cytokine genes, are able to differentially modulate the production of inflammatory cytokines and influence the risk of gastric cancer development among populations [55,56]. Although diagnostic and therapeutic advances have increased survival, the prognosis remains low mainly due to GC resistance to treatment [57]. For this reason early treatment for the infection is considered the key to prevent gastric cancer [58] and the use of antibiotics (i.e., tetracycline) as chemoprevention is a highly debated topic [59,60].

Taking into consideration factors that are involved in GC development and the drugs currently used to treat the tumor and *H. pylori* infection, we use an integrated approach to analyze this tumor. The visualization of GC as part of a complex network of interacting proteins and drug targets is helpful to better manage the massive amounts of data and to improve our understanding of the impact of pharmaceuticals on this disease.

5.3. Illustrative network for gastric cancer

There is extensive literature on gastric cancer and *H. pylori*, as demonstrated by the 7338 papers returned from the search "gastric cancer and *Helicobacter pylori*" in PubMed (http://www.ncbi.nlm.nih.gov/pubmed/). We downloaded data from select papers on the subject, where gene expression [61], methylation [62], copy number variation [63], microRNAs [64] and SNPs [65] were profiled in *H. pylori* derived gastric cancer. We also included data from two papers that conducted a network analysis [66], and a pathway analysis [66] on the same type of tumor. In addition, we collected information for drugs currently used to treat gastric cancer (from http://www.cancer.gov/cancertopics/druginfo/stomachcancer, accessed November 2013) and to target *H. pylori* [67].

Since each data set collected had different formats and identifiers, we converted individual gene identifiers to UniProt IDs using DAVID Ver. 6.7 (http://david.abcc.ncifcrf.gov/). We used the IDs to construct the network using NAViGaTOR Ver. 2.3 [32,68,69] (http://ophid.utoronto.ca/navigator), retrieving all known protein–protein interactions between them from I2D Ver. 2.1 [14,15] (http://ophid.utoronto.ca/i2d). We then added microRNAs, linked to their targets as described in the paper using mirDIP Ver. 1.1.2 [70] (http://ophid.utoronto.ca/mirDIP), and drugs, linked to their known targets as collected from DrugBank Ver. 3.0. [11] (http://www.drugbank.ca/).

The resulting network is shown in Fig. 2, and it illustrates key players in gastric cancer derived from *H. pylori* infection, as highlighted by the integration of different datasets. We also performed an analysis of all possible shortest paths and highlighted nodes with a central role. For example, one of the genes included in more than one dataset is EGFR, which is characterized by a high degree and high centrality (i.e., high number of shortest paths running through it). EGFR is a member of a family of cell surface receptor tyrosine kinases (ErbB), and plays a crucial role in development and progression of various human diseases. In particular, EGFR has been demonstrated to phosphorylate and regulate several cellular proteins and to initiate signal transduction cascades, involved in cell proliferation, migration, invasion, metastasis, angiogenesis and inhibition of apoptosis. Recent studies demonstrate the advantage of chemotherapy combined with a monoclonal antibody, trastuzumab, directed against another receptor which is a member of the same family, HER2 [71] for advanced GC patients. In addition to HER2, EGFR may also be a promising therapeutic target in gastric cancer, as shown by EGFR expression in patients with an advanced clinical GC stage or metastasis [72]. These data support the hypothesis that EGFR may play a central role in the pathogenesis and prognosis of GC. The involvement of EGFR signaling has a crucial role in development and progression of various human tumors. Nevertheless the prognostic role of EGFR in gastric cancer remains controversial [73,74], EGFR inhibitors (cetuximab) for metastatic gastric cancer are currently under investigation [75], and could improve status in at least a subgroup of GC patient. Among the several pathways involved in GC pathogenesis, immune system and inflammatory response play a major role, as demonstrated by experimental data on animal models of *H. pylori* gastritis and highlighted by our network analysis as well.

5.4. Investigating the relationship between tetracycline and gastric cancer through network analysis

Tetracyclines are antibiotics commonly used against *H. pylori* infection [76], responsible for the release of inflammatory mediators (TNFα, IL1β, IL6) by the host during infection. Epidemiological data describe the inflammatory process to be involved in aberrant DNA methylation during carcinogenesis [77]. Interestingly, tetracyclines have anti-inflammatory activity, and may reduce the risk of GC through anti-inflammatory effects that are independent of their antimicrobial activity [78]. We investigated the relationship between tetracyclines and GC by applying 'flow analysis' to a human protein–protein interaction (PPI) network.

PPI networks can provide a better understanding of drug mechanism of action by linking drug targets with biological pathways and known disease proteins. Applying flow analysis to a PPI network helps identify specific proteins and interactions that are likely modulated by a drug. Flow analysis has been previously applied to biological networks to identify lethal proteins [79], predict drug targets [80], and model gene regulatory networks [81]. In our study, the human interactome downloaded from I2D Ver. 2.1, was analyzed as an electrical circuit – interactions corresponding to resistors, and the flow of biological signals modeled as electrical current. Signals were propagated from protein targets of tetracycline, obtained from DrugBank, to protein products of genes associated with GC [82]. All interactions in the network were assigned a resistance of one. The flow algorithm considered all paths in the networks between the start and end points, and calculated

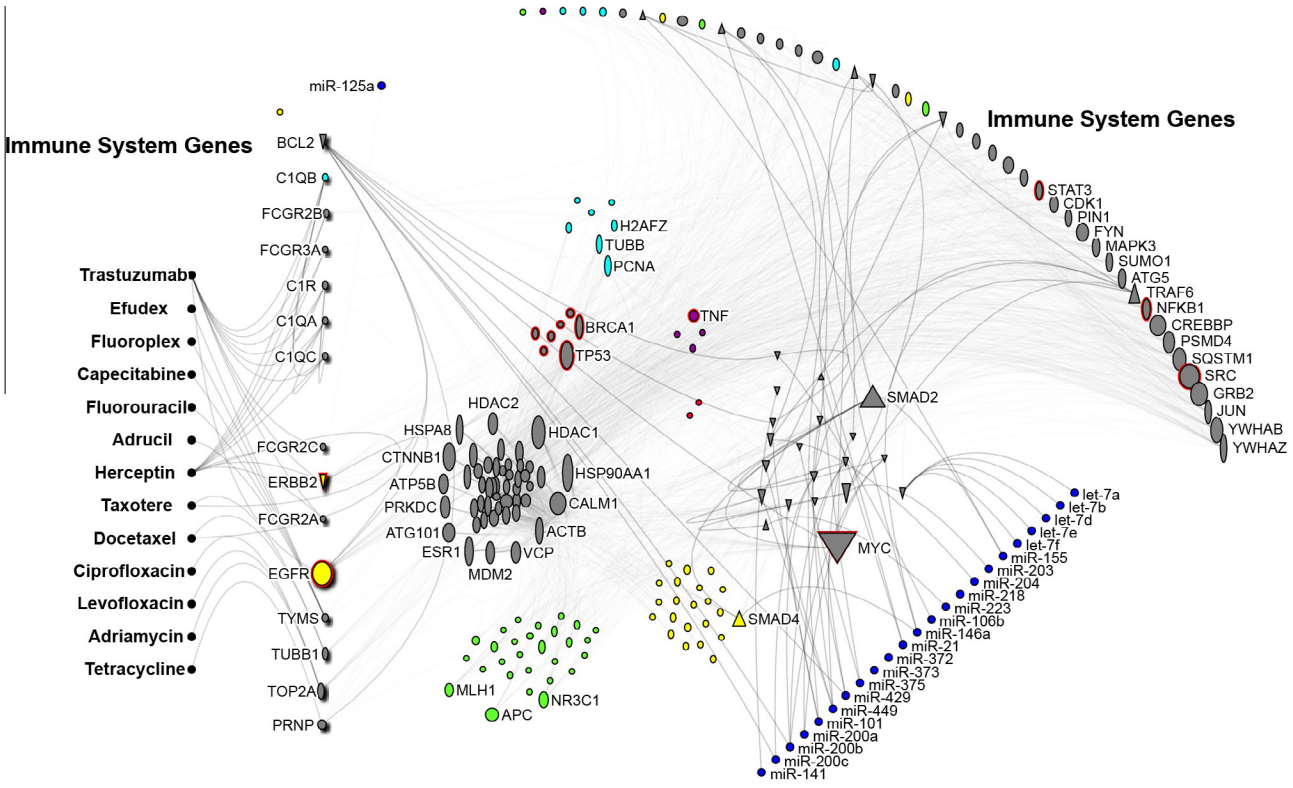

Fig. 2. Illustrative network. Network integrating different datasets from the literature for gastric cancer and *H. pylori*. Node colours: red, gene expression; yellow, copy number variation; green, methylation; purple, SNPs; turquoise, network hubs; red highlight, key pathways; blue, microRNAs; black, drugs. Upwards triangles are targets of up-regulated microRNAs, downwards triangles are targets of down-regulated microRNA. Width is proportional to all-pairs-of-shortest-paths measure. Height is proportional to degree. Upper part of the figure: Immune system related genes (pathway most enriched in our network; $p = 7.02e^{-17}$).

scores for all interactions, indicating their importance for the flow of signals. Each protein in the network was assigned the sum of the scores of its interactions; thus, proteins with high scores were considered the ones most strongly affected by tetracyclines. Our analysis revealed that tetracyclines affect the same proteins that are regulated via microRNAs, in response to *H. pylori* infection [83]. Scores of these proteins were significantly higher than others in the PPI network (Mann–Whitney U *P*-value = 6.4e − 10). The resulting network is shown in Fig. 3, constructed from the illustrative network (Fig. 2) and considering only shortest paths connections between drug targets–microRNA targets–gastric cancer associated genes. Moreover, tetracyclines could moderate inflammatory response, reducing cytokine release by neutrophils and macrophages, in particular TNFα and IL1β [84], that are present in our network as important nodes of the connection among the different elements involved in GC pathogenesis. These results provide supportive molecular evidence for the use of tetracyclines in *H. pylori*-infected patients, not only to eradicate the bacterium, but also for their additional anti-inflammatory effects.

This shows that comprehensive data integration can identify rational treatment options for patient subgroups, and network modeling provides a useful strategy to fathom drug mechanism of action.

6. Network structure–function relationship

The study of network structure has continued to be an active research area in many domains. Many real-world data, can be effectively modeled as networks. In this section, we focus on network structure analyses in biological networks, and link it to the function of these networks, components that form it, and the complete biological system.

One important area of biological network research is comparative network structure analyses. Despite the noise in current biological data, network structure analyses bring forth relevant biological insights such as lethality [85,86], protein functions [86,87], properties of disease genes [88], and drug toxicity [89]. Several methods have been developed to compare different conditions, such as differential gene expression approaches (e.g., [90–92]). Identifying individual genes that are differentially expressed between different conditions e.g., normal versus disease tissues, is useful. This approach has been widely used in deriving molecular signatures, gaining biological understanding, and hypothesizing new diagnostic as well as prognostic markers [93]. However, it has been shown that lists of differentially expressed genes across studies have low reproducibility (e.g., [94]). Furthermore, useful prognostic signatures are not necessarily the most differentially expressed genes [95]. Evidence is growing to show that network-based approaches provide substantial benefits (e.g., [96,97]). Top-scoring sub-networks are shown to overlap well with known regulatory mechanisms [98]. Sub-graphs are more reproducible, and are better predictor for breast cancer metastasis than individual genes [96]. Sub-networks are shown to be effective biomarkers in aging prediction [97]. Comparing network structures between different networks can provide novel insights. For example, comparing network structures between context-specific networks, condition-specific networks, time-specific networks provides unique structural relationships to different contexts, conditions, phenotypes, and spatio–temporal aspects.

6.1. Network properties

In order to compare and characterize complex networks, objective measures of network structure are needed. There are two main

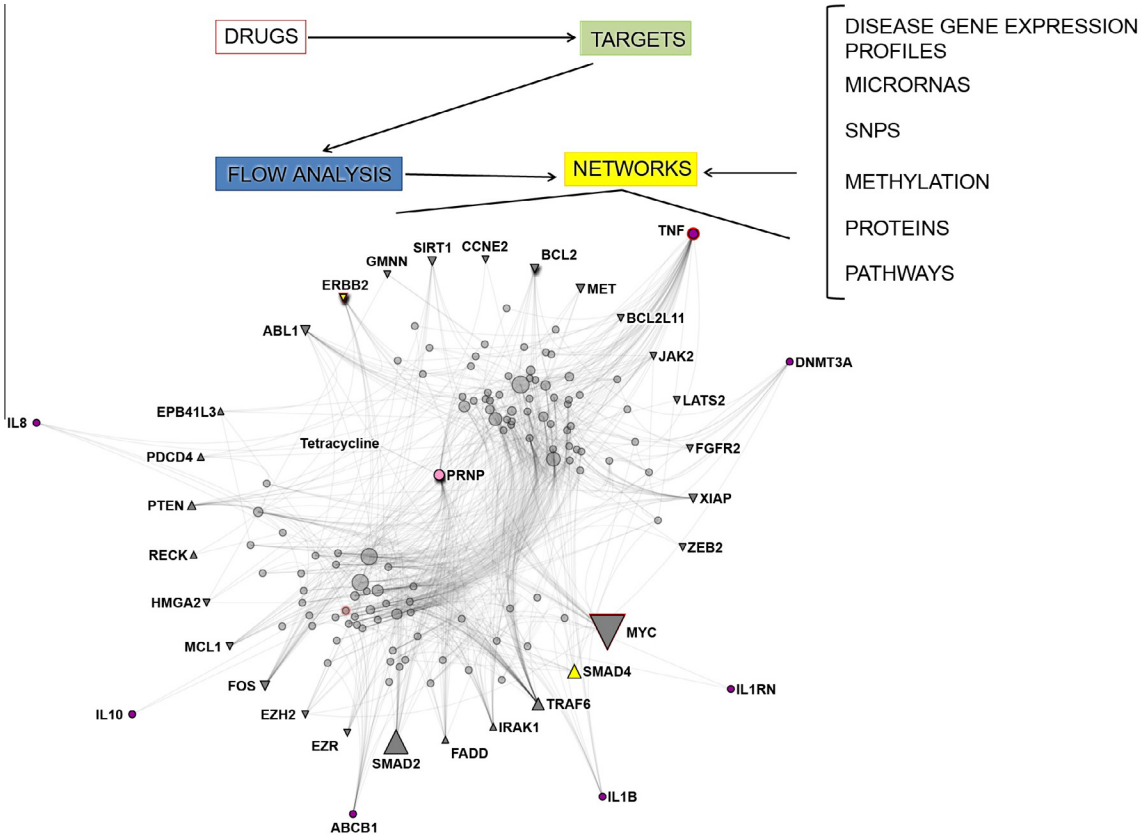

Fig. 3. Flow network. The work flow includes drugs and their known targets as collected from DrugBank (http://www.drugbank.ca/). The nodes of the network were generated by integrating data about the GC profile including gene expression studies, methylation profile, copy number variation, microRNAs, SNPs, protein and pathways involved in the disease. Flow analysis applied to a PPI network helps to identify specific proteins and interactions that might be modulated by a drug. Network based on flow from tetracyclines to gastric cancer genes and highlighting the connection to genes targeted by microRNAs. Node colours: yellow, copy number variation; purple, SNPs; red highlight, key pathways; pink, direct drug target; black, drugs. Upwards triangles are targets of up-regulated microRNAs, downwards triangles are targets of down-regulated microRNA. Node size is proportional to centrality, as measured by all-pairs-shortest-path [184].

categories of network measures for biological network comparisons, global network properties and local network properties. Global network properties examine the overall network, while local network properties focus on local patterns of the network [99].

Basic definitions used in this section are presented. $G(V, E)$ is a graph (network) where V is the set of vertices and E is the set of undirected edges [100]. $u,\ v \in V(G)$ are adjacent if there is an edge between u and v, and u is a *neighbor* of v. A *path* in a graph contained vertices that are ordered such that two vertices are consecutive in the ordering if they are adjacent. A shortest path between $u,\ v \in V(G)$ is a path with the minimum length between u, v, and the length of the shortest path between u and v is denoted as $d(u,v)$.

H is a *sub-graph* of G if $V(H) \subseteq V(G)$, $E(H) \subseteq E(G)$ and the endpoints assignment of edges in H is the same as in G. H is an *induced sub-graph* of G if H is a sub-graph such that $E(H)$ consists of all edges that are connected to $V(H)$ in G. A graph is *connected* if there exists a path between $u,\ v\ \forall\ u,\ v \in V(G)$, otherwise, the graph is *disconnected*. A graph is *complete* (also known as a clique) if there exists an edge between $u,\ v\ \forall\ u,\ v \in V(G)$.

6.1.1. Global network properties

Global network properties that are often used in biological networks include diameter, clustering coefficient, and centrality measures. The *diameter* of a graph is the maximum $d(u,v)\ \forall\ u,\ v \in V(G)$. Often in large networks analysis, a *diameter* is the average $d(u,v)\ \forall\ u,\ v \in V(G)$ [101]. The *clustering coefficient*, C, of the network is the average of C_v for all $v \in V(G)$ in the network, where C_v is the clustering coefficient for node v [102]. C_v is defined to be $\frac{2E_v}{n_v(n_v-1)}$ where E_v is the number of edges between all the neighbors of v, and n_v is the number of neighbors of v. The *clustering coefficient* measures how connected are the neighbors of any node. Several centrality measures have been applied to biological networks, and two of them are the *degree centrality* and the *betweenness centrality*. The *degree centrality* measures the importance of the role vertex u plays in a graph by measuring the number of interactions u is involved in. Let the degree of u be denoted as $d(u)$, $d(u) = \sum_{i \in V(G)} e_{ui}$, where $e_{ui} \in E(G)$. The *degree centrality* of vertex u is defined as $C_d(u) = d(u)$ [103]. The *betweenness centrality* measures the importance of vertex u in a graph by measuring the proportion of paths between other vertices in G that u is involved in [103,104]. The *betweenness centrality* of w is $BC(w) = \sum_{u,v \in V} \frac{S_{uv}(w)}{S_{uv}}$ where S_{uv} is the number of shortest paths between u and v, and $S_{uv}(w)$ is the number of shortest paths between u and v that pass through w, and $\{u, v, w \in V \mid u \neq v, v \neq w\}$.

6.1.2. Local network properties

Global network properties examine network properties of entire networks, but more detailed network properties are needed for comparing different networks. In this section, we discuss local network properties, motifs and graphlets. *Network motifs* [105] are small sub-graphs in a network such that when compared to randomized networks, their structures appear significantly more. Different motifs are found in different complex networks, for example, the feed-forward loop, a 3-node motif, is found in

transcription-regulatory networks, but the feed-forward loop motif is under-represented in food webs networks [105]. *Graphlets* are all non-isomorphic connected induced graphs on a certain number of vertices [106], and by definition, they have the ability to capture all the local structures on a certain number of vertices. There are differences between graphlets and motifs. Motifs are dependent on the randomization scheme, but graphlets are not as they do not have to be over-represented when compared to randomized networks.

Several graphlet-based network properties have developed, and one of them is the *relative graphlet frequency distance* (*RGF-distance*) [106]. RGF-distance has been used to determine which random graph model is the most accurate representation of PPI networks [106]. RGF-distance measures the distance between two networks using graphlet frequencies – a count of the number of graphlets of each type (from 1 to 29) in a network. $N_i(G)$ is the number of graphlets of type i, $i \epsilon \{1,\ldots,29\}$ in graph G, and $T(G) = \sum_{i=1}^{29} N_i(G)$ is the total number of graphlets in graph G. The *relative frequency of graphlets* is defined to be $\frac{N_i(G)}{T(G)}$, and the *relative graphlet frequency distance* between graphs G and H is $D(G,H) = \sum_{i=1}^{29} |F_i(G) - F_i(H)|$, where $F_i(G) = -\log \frac{N_i(G)}{T(G)}$.

6.2. Computational challenges

Large network comparisons are challenging as comparing all aspects of networks involves the sub-graph isomorphism problem which is NP-complete [107]. Given two graphs G and H as input, the sub-graph isomorphism problem is to determine if there exists a sub-graph in G such that it is isomorphic to H.

Since the sub-graph isomorphism problem is computationally intensive, a reasonable method is to restrict the sub-graph size for searching (e.g., [108,109]). Another method is to approximate sub-graph frequencies instead of counting the exact sub-graph frequencies (e.g., [110,111]). Furthermore, two heuristics to efficiently estimate graphlet frequency distribution for high confidence PPI networks and geometric random networks have developed, *Neighborhood Local search* (*NLS*) and *Targeted Node Processing* (*TNP*) [112]. The NLS heuristic searches for a specific graphlet in the neighborhood of a randomly chosen vertex, and the TNP heuristic reduces search space for graphlet enumeration. The idea of TNP is to have an exhaustive search for graphlets in a small part of the network, and the graphlet frequency distribution obtained from the exhaustive search is used to estimate the graphlet frequency distribution of the entire network.

7. Importance of the interactome in network analysis

7.1. Expansion of the known human interactome

From 2003 to 2013, the number of known human interactions has increased more than 3-fold. The known interactome has become an important resource for applications such as prediction of protein function [113], prediction of disease genes [114], and multiple aspects of drug development [115]. However, the rapid growth of the known interactome comes with caveats: the network is influenced by biases and the accuracy of many interactions is in question. Here, we briefly review how the interactome has expanded, the accompanying benefits and continuing challenges. We focus on the interactome in the I2D database.

7.2. Interactome expansion

The known human interactome began in the 1960s (Fig. 4). Interactions were identified by methods such as co-immunoprecipitation and affinity chromatography, often used in combination. Individual studies using small-scale screens, usually identified less than 20 interactions but by the year 2000, 13,363 interactions had been identified through 11,390 studies. From about 2002, HT studies began to identify human PPIs. Individual studies identified up to 6463 PPIs using methods such as HT yeast 2-hybrid, protein arrays, and HT mass spectrometry-based methods {Ewing, 2007 #32108}. From 2001 to 2013 the number of interactions increased more than 5 times, with over 56,000 coming from HT studies. Since mass spectrometry-based methods report protein complexes, the number of binary PPIs they contribute can be difficult to assess. Typically, complexes are converted to binary interactions using a spoke model, where the bait interacts with all complex members, which can be an overestimation.

7.3. Benefits of interactome expansion

The completeness of the known interactome is critically important for tasks such as prediction of disease genes and drug targets. The percentage of disease proteins and drug targets in the known interactome has increased from about 65% in 2003 to 88% in 2013 (Fig. 5). Interestingly, most of these proteins were added to the interactome through small-scale screens.

7.4. Interactome biases

The known interactome is not a random subset of human PPIs; certain types of interactions and proteins are over-represented while others are significantly deficient. This is largely a consequence of research bias and limitations of PPI detection methods. Research bias has been an important problem in proteome and interactome research; most studies focus on a relatively small number of proteins that have been previously investigated [116]. Studying such proteins is often less risky, less time consuming, and cheaper. Since the functions and disease relevance of these proteins may be partially known, it is easier to formulate hypotheses about them, design short term studies, and procure funding. In addition, experimental tools and protocols may be available for

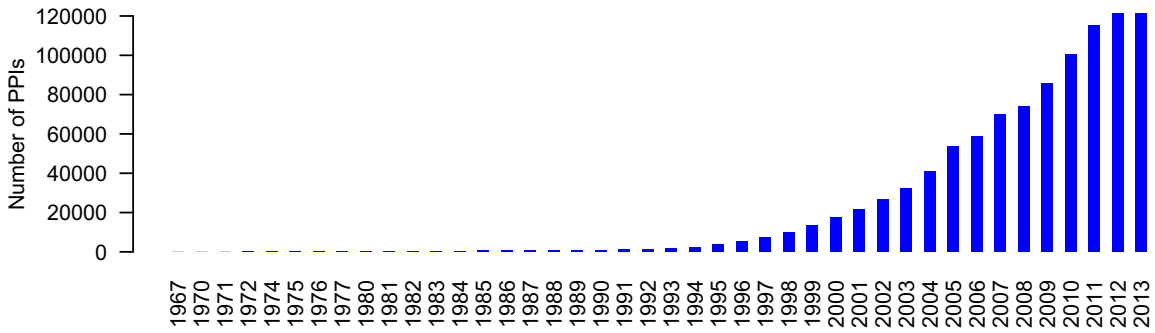

Fig. 4. Cumulative number of known PPIs by year.

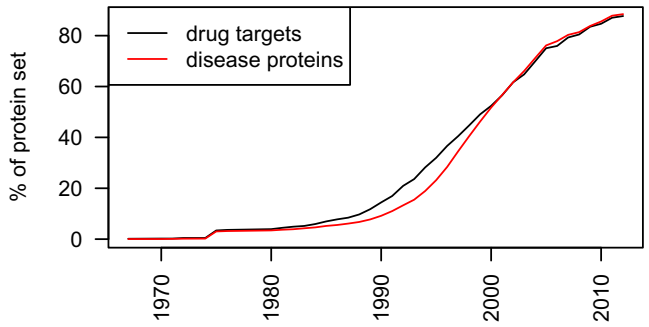

Fig. 5. Percentages of drug targets and disease proteins in the known interactome, by year. Drug targets were obtained from DrugBank, version 3.0 [11]. Disease genes were obtained from the DOLite database (http://django.nubic.northwestern.edu/fundo/download), the Genetic Association Database (GAD) [185], and the Comparative Toxicogenomics Database (CTD) [186].

studying the proteins [116]. Consequently, research bias has had a profound impact on the growth of the interactome. New interactions have typically been identified between proteins that already have known interactions (Fig. 6); about 1/3rd of human proteins remain absent from the known interactome – resulting in interactome orphans. The number of new proteins being added to the interactome through small scale screens has stopped increasing since about 2003 (Fig. 7).

While research bias has been primarily a problem for small-scale studies, HT screens have their own biases. They tend to detect interactions involving proteins that are older, more conserved across species and highly expressed; other types of interactions may be poorly detected [117]. For example, Tandem Affinity Purification is less efficient for transient interactions, or interactions involving short or low abundance proteins [118]. HT studies have integrated many new proteins into the known interactome, especially in 2005, but their contributions have not been increasing (Fig. 7).

Biases in the known interactome have led to controversy in the interpretation of network structure [119]. Many studies have drawn biological conclusions from topological analysis of network structure. For example, central nodes have been associated with essentiality [85], and the observed scale-free structure of some networks has been interpreted as a strategy that confers robustness to PPI networks [120]. However, several studies have shown that these properties of PPI networks may be due to biases [121].

7.5. Interactome accuracy

Assessing the accuracy of PPI networks has been an ongoing challenge, as most evaluation strategies are subject to their own biases. Several studies evaluated PPIs based on assumptions about the properties of interacting protein pairs: shared localization and function, and correlation of encoding gene pairs [122,123]. However, localization and function information is less effective if proteins are poorly annotated, or tend to migrate between cellular compartments. Correlation of encoding gene pairs does not hold for transiently interacting proteins [124]. Some studies have evaluated PPIs based on reference sets comprising trusted interactions; typically interactions detected multiple times. However, such interactions usually comprise proteins that are frequently studied [119], and are devoid of interactome orphans.

7.6. Comparisons of human interactomes across time

As discussed above, the size of human interactomes has grown enormously over the last decade. The growth from 1967 to 2013 can be compared using graphlet distributions, and data from I2D. As mentioned before, graphlets are all non-isomorphic connected induced graphs on a certain number of vertices [106]; they have the ability to capture local structures. Fig. 1 in [106] reviews all 29 3- to 5-node graphlets. Graphlet numbers increase monotonically with the number of edges between the 5 vertices. For each

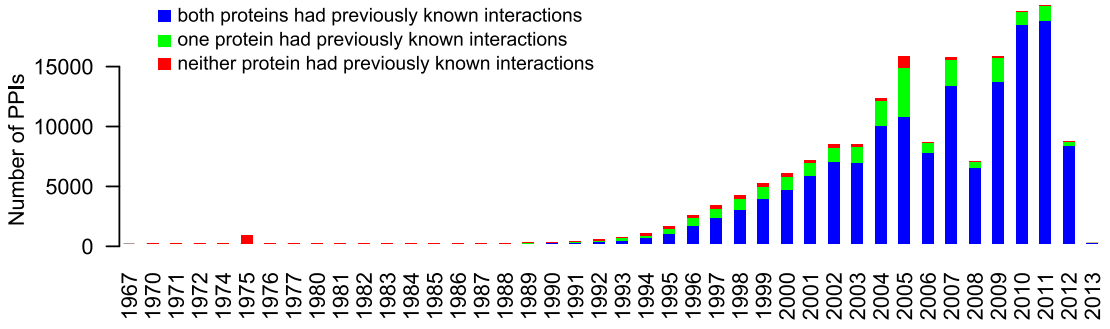

Fig. 6. Numbers of PPIs added to the known interactome each year.

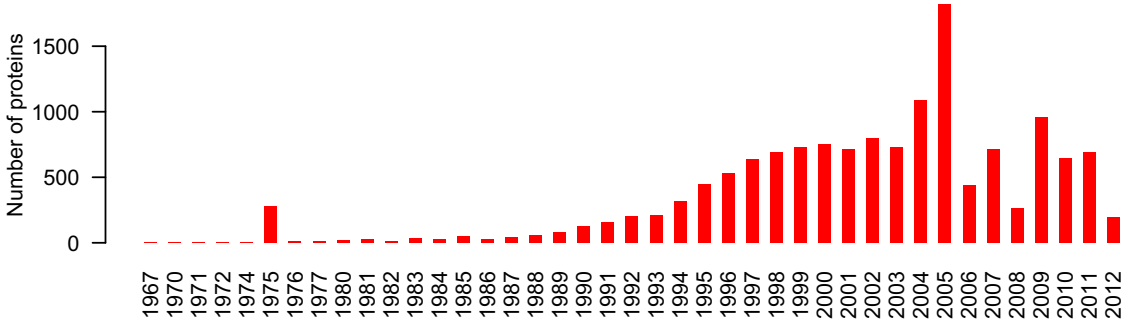

Fig. 7. Numbers of proteins added integrated into the known interactome each year.

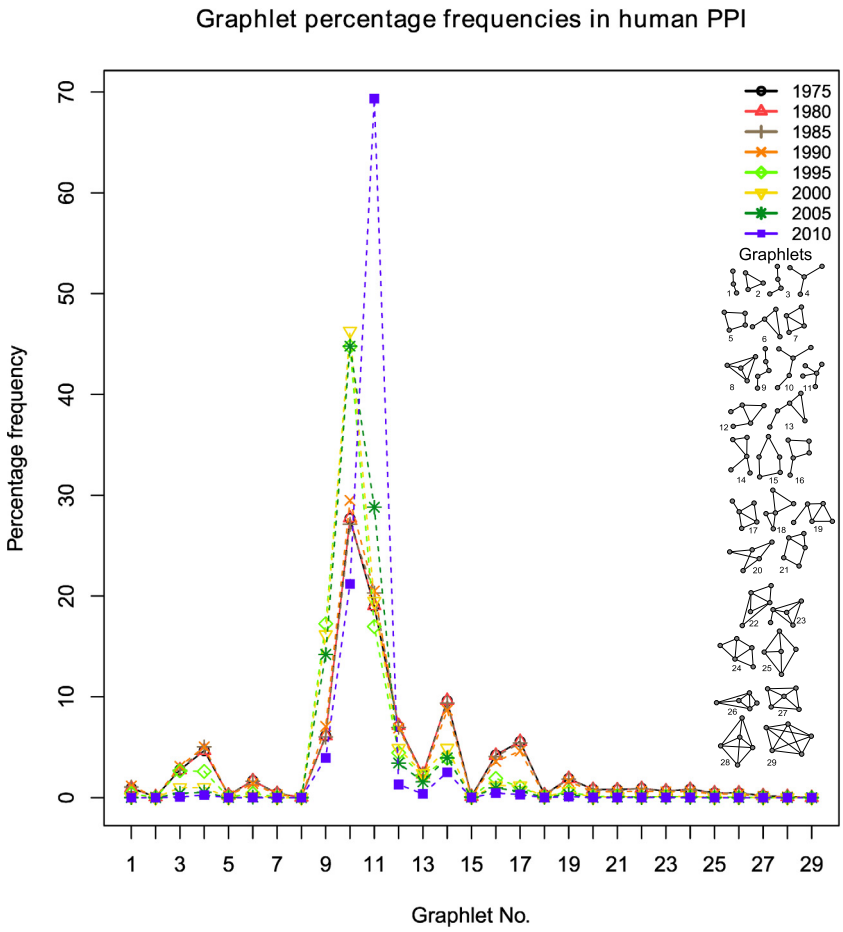

Fig. 8. Graphlet percentage frequencies in human PPI every 5 years, from 1975 to 2010.

human PPI network, G, a count of the number of graphlets of each type (1 to 29) is computed using GraphCrunch 2 [125]. $N_i(G)$ is the number of graphlets of type $i, i \epsilon \{1, \ldots, 29\}$ in graph G, and $T(G) = \sum_{i=1}^{29} N_i(G)$ is the total number of graphlets in graph G. The graphlet percentage frequency is $\frac{N_i(G)}{T(G)} \times 100$.

Fig. 8 shows graphlet percentage frequencies for every 5 year, from 1975 to 2010, and Fig. 9 shows graphlet percentage frequencies for the last 5 years. The graphlet percentage frequency distributions for 2010 to 2013 are fairly similar, but the distributions are different from 1975 to 2010.

In Fig. 8, graphlet No. 9 is substantially more frequent in 1995–2005 than 1975–1990. Graphlet No. 9 is a linear line, and there is a huge increase in both the number of proteins and the number of PPIs (Figs. 6 and 7) in 1995–2005 when compared to 1975–1990; thus, the increase of graphlet No. 9. 2010 has the lowest percentage of graphlet No. 9 which is not surprising as from 2006 to 2010, most edges that are added to the graphs consist of vertices that are already in the graphs (Fig. 6). A similar trend is observed for graphlet No. 10, which resembles graphlet No. 9 – linear line with a "v" shape at one end.

In Fig. 8, the peak is at graphlet No. 11 for 2010, a huge increase from 2005. Graphlet No. 11 accounts for 69.36% of graphlets in 2010, and this trend continues for years 2010–2013 (Fig. 9). Graphlet No. 11 has 5 nodes and 4 edges, the same number of nodes and edges as graphlet No. 9 and 10. The difference is that one node in graphlet No. 11 has degree 4, which connects to all other nodes in the graphlet. Many factors can lead to the huge increase of graphlet No. 11 from 2005 to 2010. For example, trivially, graphlet No. 11 can be formed by adding edges to the graph such that one of their vertices is already in the graph, but the other vertex is not. Graphlet No. 11 can also be formed by adding edges to the graph such that both of their vertices are already in the graph. For example, edge e consists of vertices a and b, a and b are already in the graph but belong to different components; graphlet No. 11 can be created by adding e, which connects different components in the graph. From 2006 to 2010, there are many edges that are added such that both of their vertices are already in the graph (Fig. 6). Interestingly, there are 257 connected components in 2005 and 145 components in 2010. Furthermore, the size of the giant component increases; from having 10,258 nodes and 51,065 edges in 2005 to having 13,462 nodes and 97,637 edges in 2010. This analysis suggests that recently the tendency is to study new interactions between old proteins, especially focusing on proteins already well-studied. It also demonstrates the bias in protein interaction studies and suggests that we are far from improving our knowledge of the role and function of interactome orphans (Fig. 6).

The decrease in graphlet No. 9 and graphlet No. 10 for 2010 is not surprising as the number of edge to node ratio increases from 1.24 to 7.38 from 1975 to 2010. Due to the change in the edge to node ratio from 1975 to 2010, a natural question to ask is: are there higher percentages of denser graphlets in 2010 than in earlier years? Fig. 8 shows that this is not the case. 2010 has low graphlet percentage frequencies for dense graphlets as the edge to node ratio is not high enough for a high percentage of graphlets to be dense. Interestingly, there are higher graphlet percentage frequencies for graphlet No. 14, 16, and 17 in 1975–1990 than in 1995–2010. In the early years, experiments were focused around

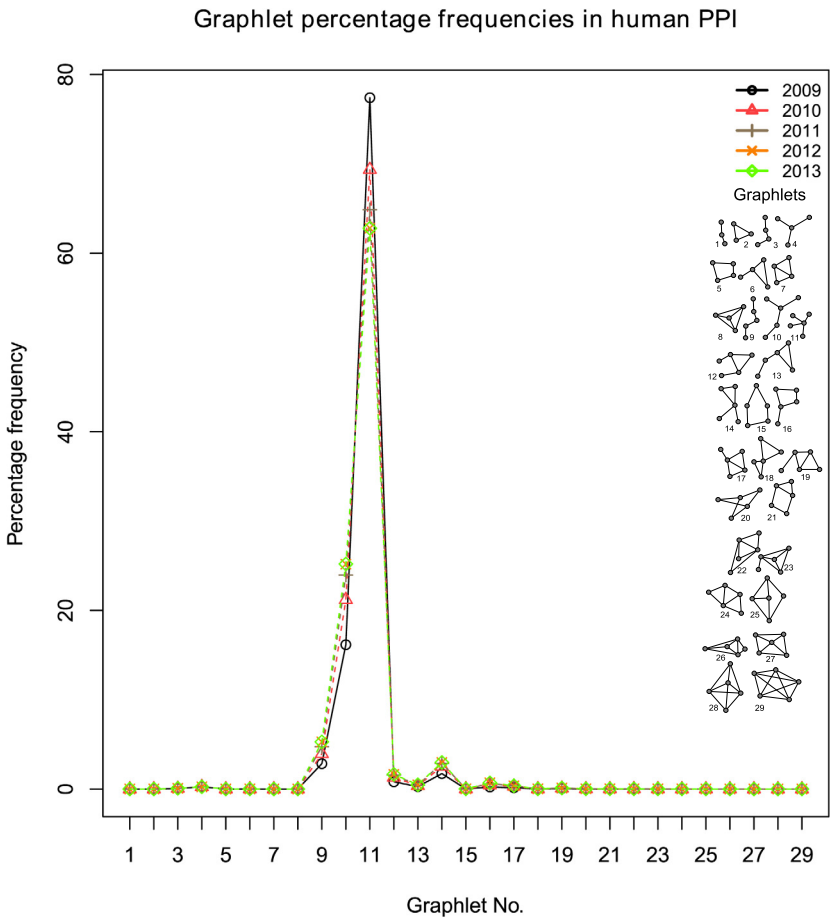

Fig. 9. Graphlet percentage frequencies in human PPI over the last 5 years.

a small number of proteins that were of interest resulting in some denser regions.

There are some graphlets that have low percentages throughout all these years. For example, graphlet No. 12 to 16 all have 5 nodes and 5 edges; the percentage graphlet frequency for graphlet 15 is low from 1975 to 2010, but this is not the case for graphlet No. 14 or 16. The observation suggests that some graphlets are rarer in PPI networks. Graphlet No. 15 is a ring with 5 nodes and 5 edges. For example, it is not likely for a lot of graphlet No. 15 to be present in signaling pathways as signaling pathways usually have more redundancy between proteins than having just a ring between them.

8. Power of network analysis to identify context-specific information

8.1. Context-specificity of interactomes

While the current interactome maps serve as a rich resource in systems and molecular biology research, interpreting them is challenging partly due to the lack of information about the dynamics or 'context'-specificity of the PPIs. The notion of 'context' can cover a wide range of details about the conditions that allow the PPI to form, e.g., organism, tissue, cell type, disease stage, cellular compartment, biological process, time point, etc. However, global PPI maps in their present form lack this information as they are the result of combining the PPIs that have been captured in artificial HT settings (e.g., yeast-two-hybrid systems) or manual curation of the low-throughput studies without including annotations about the conditions of the experiments. In the past decade we witnessed increased interest in knowledge about the context of the PPIs.

While widely used low-throughput methods still serve as an important resource for uncovering detailed PPIs, their context, and their mechanisms [126,127], developing HT systems for screening context-specific PPIs are growing fast [128–131]. In an earlier effort in applying HT screens to identify function-specific PPIs, Lehner and Sanderson have applied yeast-two-hybrid screens to identify PPIs and complexes that contribute to mRNA deregulation [132]. In more recent efforts, Behrends et al. used mass spectrometry to extract a network of autophagy PPIs [133] (process specificity), Hegele et al. applied yeast-two-hybrid to identify PPIs specific to the spliceosome [134] (cellular compartment specificity), and Li et al. combined mass spectrometry and machine learning to describe the network of EGFR interactions and its core in different EGFR-mutated lung cancer cell lines [135] (tissue/cell type specificity). Although as soon as these PPIs get published, data curators for global PPI maps include them into the global networks, no major effort so far has been made to publish a compiled library of the context-specificity of such PPIs.

The second class of approaches for adding context to the PPI networks includes computational methods that predict the context by systematic integration of different sources of biological information. The first premise behind most of these methods is that in order for a PPI to form, the two proteins need to be present in the same environment [96,136]. Besides, knowledge about the RNA-sequences of the interacting proteins and functional relevance between protein pairs have been used to predict the context of the PPIs [130,131,137].

Table 1
Resources for context-specific PPIs.

Resource	Context	Reference	Data types (sources)	# Of covered PPIs	Data availability	Web service
Bossy and Lehner (2009)	Tissue	PMID: 19357639	PPIs, BioGPS,	80,923 (Experimental and interologs)	Tab delimited file through journal	NA
TissueNet	Tissue	PMID: 23193266	PPIs, BioGPS, ProteinAtlas, RNA-seq data	67,439 (11,225 Proteins)	Only web query	Query
HIPPIE	Tissue, subcellular locations, biological processes	PMID: 23300433	PPIs, BioGPS, GO	>97,000 Annotated PPIs	Tab delimited (only PPIs and their sources)	Query (tissue-specificity)
HCPIN	Cancer (structure-function information for the human cancer protein interactome)	PMID: 18487680	Pathways, physical PPIs, Cancer Gene Census Database	9784 (2978 Proteins)	Only web query	Query
cisPIN	Cancer (interface information about interactions of cancer proteins)	PMID: 20011507	PPIs, PDB	CSPIN: 3221 (1303 proteins)	NA	NA

Gene expression datasets are the most widely used resource whose integration with PPI networks has resulted in prediction of context-specific PPI networks. This approach has shown promising results in extracting tissue-specific PPI networks and analysing their properties [136,138–142]. Many of such works use gene expression data from BioGPS [143]. Also prediction of cell type (e.g., healthy versus disease cells) specific interactomes through this strategy combined with graph mining algorithms has shown potential in improving our understanding of molecular modules behind cellular processes and functions [97] as well as disease classification and prediction of disease outcomes [96]. While these efforts show the increased interest and improvements in identifying or inferring context-specificity of the PPIs, most of these methods have been applied to only one or a few conditions. More recently some groups have initiated compiling larger sets of predicted context-specific PPIs whose summary is shown in Table 1.

8.2. A context application: modeling of signaling pathways

Modern HT and multiplex experimental technologies have enabled near-complete identification of the total 'parts' compendium of the bio-molecular signaling machinery. Coupled with innovations in high-speed and high-content microscopy and as well, with advancements in image-processing, expansive vistas in information space have opened that may shed new light on how signaling mechanisms govern biological and pathological responses [144,145]. However, this progress has come at the expense of significant new challenges in data analyses and management and perhaps most significantly, in data interpretation [146]. The sheer density and scale of detailed information at the level of individual 'parts' has stressed a need for understanding function at the level of pathways and networks and more importantly, to identify their relation to phenotypic or pathological endpoints. Uncovering the functional basis of pathways and networks will be essential for achieving the therapeutic goals of medicine and drug discovery.

The need to understand signaling at the level of pathways and networks has motivated the study of large-scale PPI networks that are generally broad in scope and highlight influences and connections between nodes in the system of interest, but their abstract nature, especially at initial assembly stages, provides limited mechanistic and functional insight. In attempts to increase functional insights, networks specific to cell types or states have been created as mentioned before. Some of these strategies were highlighted in the development of a 263-node mutant EGFR interactome in lung cancer [135] starting from a subset of 102 networked proteins on the basis of their annotation to the ErbB, MAPK, apoptotic and cell-cycle pathways [135]. A functional lethality screen using siRNA directed against this 102-protein subset narrowed it to a core set of 14 survival-associated proteins, which when associated with drug networks, led to the identification of anti-resistance compounds. Although the network model was initially enabled by use of large MS-based interaction data sets, functional and therapeutic insights were facilitated by prior knowledge in a specific biological system of interest. Similarly, a novel PPI screen (MaMTH) has led to identification of novel L858R EGFR mutant-specific interactions, opening new possibilities for understanding pharmacological response to Erlotinib.

Other approaches in the study of cell signaling have involved so-called 'data-driven' modeling approaches, which depend on input from information-dense signaling data sets [147]. Data-driven models employ statistical inference techniques such as clustering, principle component analyses, and partial least squares regression to predictively relate signaling state(s) to various phenotypic endpoints [147–149]. Hence, inference modeling approaches provide greater functional insights compared to network models derived from single- or multi-dimensional interaction data, albeit at the expense of network size. The applicability and utility of inference modeling is highlighted by a study in which partial least square analysis of data sets harboring 7980 intracellular signals and 1440 cell responses identified signaling metrics predictive of cytokine-induced apoptosis [148]. Inference methods have also been used to directly assemble network models using functional signaling data sets generated by proteomic technologies [150,151]. One such study used primary hepatocytes and hepatocellular carcinoma cell systems to show differences in network topology representative of differences in inflammatory signaling between the systems [151]. Multi-linear regression analyses of approximately 26,000 protein-state measurements derived from growth factor/cytokine signaling experiments was used to infer the differences in networks. The latter approach represents an alternative biochemically-focused approach to network construction capable of providing bona-fide functional insight. Inference modeling approaches have now emerged as a standard approach in system-based analysis of signaling networks [147]. Their principle strengths lie in providing strong functional insights through correlations between the dynamic state of the signaling network (as discrete signal metrics) and a particular cell state or response.

Knowledge-based physiochemical models (PM) are commonly used for studying the dynamical behavior of signaling pathways and networks. PM models use established physiochemical theory to represent signal changes (i.e., phosphorylation and protein interactions) as sets of ordinary differential equations [152]. They rely strongly on prior biochemical knowledge and require extensive calibration by quantitative kinetic data sets [152]. PM models have been termed mechanistic models because they

capture higher-levels of mechanistic detail by explicitly simulating the dynamical continuum of biochemical interactions and post-translational modifications taking place in space and time. Simulations have revealed emergent behaviors or functions of the system not apparent from topology alone [153]. PM models therefore harbor significantly better mechanistic and predictive capacity relative to network or inference models, and have been broadly employed for studying the temporal mechanisms by which cellular signal systems operate at the pathway and network level [154–165].

Currently, PM/mechanistic models are constrained in scale for a number of reasons; these include: (1) availability of time-resolved biochemical and genetic data; (2) the requirement for resource- and time-intensive, experimentally-integrated iterative development processes; (3) combinatorial explosions in the number of distinct species and species-states generated during pathway and network simulation [166]; and (4) a requirement for extensive prior biochemical knowledge. Development of HT MS-based phosphoproteomics [167], protein microarray technologies [168] and in cell immunocytochemical [149] approaches are making inroads towards generating larger quantities of kinetic data sets and integrating them efficiently into iterative systems-level development processes [169].

Similarly, the recent development of rule-based modeling approaches hold promise for dealing with the combinatorial explosions of molecular states possible during signal transduction [166,169,170]. These factors may propel PM models towards bringing their full capacity for mechanistic insight and predictive potential closer to omic scales of complexity in the not too distant future [152]. By way of comparison, static network modeling approaches already incorporate data on omic scales, and arguably, have almost reached their threshold capacities of prediction and mechanistic insight [152]. In time, PM models have been envisioned to realistically simulate signaling behavior, to make significant contributions to multi-component therapy development and to become economical and efficient components of preclinical testing [164,165,171,172]. In the interim, transitional modeling frameworks have been proposed. Boolean modeling approaches, which have previously been used to study metabolic and genetic regulatory networks [173–175], constitute one potential approach by which to bridge the gap between static network and PM signaling models [176–180].

Integrated modeling approaches have also been proposed in which new biological relations are inferred from large-scale omic-level networks and subsequently used to inform mechanistic simulations aimed at elucidating underlying drug-effect mechanisms in normal and disease pathways [181]. The latter examples mark the emergence of new efforts in network biology research involving a switch from static to dynamical network analyses. This new area of research will herald a significant increase in our understanding of the basis of cell response and behavior in normal and disease contexts [182].

9. Discussion

The development of new HT experimental approaches has been generating increasingly massive and complex genomic and proteomic data sets since 1990, when the Human Genome Project started. Ever since medical and biological research has needed a different approach to evaluate results and elaborate hypothesis. Although now researchers could obtain a more complete and accurate vision of a disease using multiple heterogeneous and distributed resources, a new challenge is to resolve the differences among these databases, including the identifiers, confidence and context. For example, knowing that two proteins interact, the question remains whether the experiment was done with constructs of the same species, whether full-length proteins or variants were used, and in which cells and under what conditions was the interaction detected. Comprehensive curation efforts address these issues systematically for all reviewed interactions [19,20,183], but more data exist in literature, beyond what manual curation can handle at this time. Some database interfaces are quite complex and not suitable for biologists that usually prefer easy to use tools. Importantly, databases that are geared towards biologists frequently do not support bioinformatics workflows and thus are not suitable for integrative computational analyses.

The development of network visualization has improved the possibility to represent complex phenomena, such as multi-genic human disease, in a more intuitive way. Moreover, computational studies have supported biomedical research improving measure of network properties, and supporting a reliable integration and comparison among different biological networks. However, considering that network analysis applied in the study of PPI uses available interactomes, other caveats remain to be overcome, such as interactome biases, orphans, and the accuracy and context of reported interactions. The sources of PPI include small- and large-scale studies often without reporting the experimental conditions, limiting in some case the robustness of PPI analysis. Data curators are involved in the effort to overcome this bias by including all information available to create a context-specific interactome resource, as reported above. In order to improve network mechanistic and functional insights integrated modeling approaches have been developed to better elucidate multifactorial disease and possible drug effects. An integrated PPI network could be useful to study multi-genic disease, where several omics datasets are integrated to generate hypotheses, as highlighted in our gastric cancer example. Moreover, applying the flow analysis to PPI network has improved the identification of specific proteins and interactions that are modulated by drugs, showing how network analysis might improve our knowledge of disease mechanism and suggest alternative strategies for treatment.

Acknowledgments

This research was supported in part by the grants from Ontario Research Fund (GL2-01-030, RE-03-020), and by the Friuli Exchange Program. Computational resources were supported in part by grants from Canada Foundation for Innovation (CFI #12301, #203373, #29272, #225404) and IBM. I.J. is Canada Research Chair. This research was funded in part by the Ontario Ministry of Health and Long Term Care. The views expressed do not necessarily reflect those of the OMOHLTC.

References

[1] B. Kuska, Beer, bethesda, and biology: how “genomics” came into being, J. Natl. Cancer Inst. 90 (1998) 93.
[2] R. Higdon, E. Kolker, A predictive model for identifying proteins by a single peptide match, Bioinformatics 23 (2007) 277–280.
[3] S. Behjati, P.S. Tarpey, What is next generation sequencing?, Arch Dis. Child. 98 (2013) 236–238.
[4] J.D. Tenenbaum, S.A. Sansone, M. Haendel, A sea of standards for omics data: sink or swim?, J Am. Med. Inform. Assoc. 21 (2013) 200–203.
[5] M. Chadeau-Hyam, G. Campanella, T. Jombart, L. Bottolo, L. Portengen, P. Vineis, B. Liquet, R.C. Vermeulen, Deciphering the complex: methodological overview of statistical models to derive OMICS-based biomarkers, Environ. Mol. Mutagen. 54 (2013) 542–557.
[6] G. Liu, M. Feng, Y. Wang, L. Wong, S.-K. Ng, T.L. Mah, E.J.D. Lee, Towards exploratory hypothesis testing and analysis, in: IEEE 27th International Conference on Data Engineering, IEEE, 2011, pp. 745–756.
[7] K. Van Deun, T.F. Wilderjans, R.A. van den Berg, A. Antoniadis, I. Van Mechelen, A flexible framework for sparse simultaneous component based data integration, BMC Bioinformatics 12 (2011) 448.
[8] T.U. Consortium, The universal protein resource (UniProt) in 2010, Nucleic Acids Res. (2010) D525–D531.
[9] M. Rebhan, V. Chalifa-Caspi, J. Prilusky, D. Lancet, GeneCards: a novel functional genomics compendium with automated data mining and query reformulation support, Bioinformatics 14 (1998) 656–664.

[10] K.D. Pruitt, T. Tatusova, D.R. Maglott, NCBI reference sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins, Nucleic Acids Res. 33 (2005) D501–D504 (Data issue).
[11] C. Knox, V. Law, T. Jewison, P. Liu, S. Ly, A. Frolkis, A. Pon, K. Banco, C. Mak, V. Neveu, Y. Djoumbou, R. Eisner, A.C. Guo, D.S. Wishart, DrugBank 3.0: a comprehensive resource for 'omics' research on drugs, Nucleic Acids Res. 39 (2011) D1035–D1041.
[12] D.S. Wishart, C. Knox, A.C. Guo, S. Shrivastava, M. Hassanali, P. Stothard, Z. Chang, J. Woolsey, DrugBank: a comprehensive resource for in silico drug discovery and exploration, Nucleic Acids Res. 34 (2006) D668–D672.
[13] S. Orchard, M. Ammari, B. Aranda, L. Breuza, L. Briganti, F. Broackes-Carter, N.H. Campbell, G. Chavali, C. Chen, N. Del-Toro, M. Duesbury, M. Dumousseau, E. Galeota, U. Hinz, M. Iannuccelli, S. Jagannathan, R. Jimenez, J. Khadake, A. Lagreid, L. Licata, R.C. Lovering, B. Meldal, A.N. Melidoni, M. Milagros, D. Peluso, L. Perfetto, P. Porras, A. Raghunath, S. Ricard-Blum, B. Roechert, A. Stutz, M. Tognolli, K. van Roey, G. Cesareni, H. Hermjakob, The MIntAct project–IntAct as a common curation platform for 11 molecular interaction databases, Nucleic Acids Res. 42 (2013) D358–D363.
[14] K.R. Brown, I. Jurisica, Online predicted human interaction database, Bioinformatics 21 (2005) 2076–2082.
[15] K.R. Brown, I. Jurisica, Unequal evolutionary conservation of human protein interactions in interologous networks, Genome Biol. 8 (2007) R95.
[16] Y. Niu, D. Otasek, I. Jurisica, Evaluation of linguistic features useful in extraction of interactions from PubMed; application to annotating known, high-throughput and predicted interactions in I2D, Bioinformatics 26 (2010) 111–119.
[17] C. Stark, B.J. Breitkreutz, T. Reguly, L. Boucher, A. Breitkreutz, M. Tyers, BioGRID: a general repository for interaction datasets, Nucleic Acids Res. 34 (2006) D535–D539.
[18] M. Kanehisa, A database for post-genome analysis, Trends Genet. 13 (1997) 375–376.
[19] S. Orchard, S. Kerrien, S. Abbani, B. Aranda, J. Bhate, S. Bidwell, A. Bridge, L. Briganti, F.S. Brinkman, G. Cesareni, A. Chatr-aryamontri, E. Chautard, C. Chen, M. Dumousseau, J. Goll, R.E. Hancock, L.I. Hannick, I. Jurisica, J. Khadake, D.J. Lynn, U. Mahadevan, L. Perfetto, A. Raghunath, S. Ricard-Blum, B. Roechert, L. Salwinski, V. Stumpflen, M. Tyers, P. Uetz, I. Xenarios, H. Hermjakob, Protein interaction data curation: the international molecular exchange (IMEx) consortium, Nat. Methods 9 (2012) 345–350.
[20] B. Aranda, H. Blankenburg, S. Kerrien, F.S. Brinkman, A. Ceol, E. Chautard, J.M. Dana, J. De Las Rivas, M. Dumousseau, E. Galeota, A. Gaulton, J. Goll, R.E. Hancock, R. Isserlin, R.C. Jimenez, J. Kerssemakers, J. Khadake, D.J. Lynn, M. Michaut, G. O'Kelly, K. Ono, S. Orchard, C. Prieto, S. Razick, O. Rigina, L. Salwinski, M. Simonovic, S. Velankar, A. Winter, G. Wu, G.D. Bader, G. Cesareni, I.M. Donaldson, D. Eisenberg, G.J. Kleywegt, J. Overington, S. Ricard-Blum, M. Tyers, M. Albrecht, H. Hermjakob, PSICQUIC and PSISCORE: accessing and scoring molecular interactions, Nat. Methods 8 (2011) 528–529.
[21] M. Ashburner, C.A. Ball, J.A. Blake, D. Botstein, H. Butler, J.M. Cherry, A.P. Davis, K. Dolinski, S.S. Dwight, J.T. Eppig, M.A. Harris, D.P. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis, J.C. Matese, J.E. Richardson, M. Ringwald, G.M. Rubin, G. Sherlock, Gene ontology: tool for the unification of biology. The gene ontology consortium, Nat. Genet. 25 (2000) 25–29.
[22] S. Orchard, H. Hermjakob, R. Apweiler, The proteomics standards initiative, Proteomics 3 (2003) 1374–1376.
[23] M. Safran, I. Dalah, J. Alexander, N. Rosen, T. Iny Stein, M. Shmoish, N. Nativ, I. Bahir, T. Doniger, H. Krug, A. Sirota-Madi, T. Olender, Y. Golan, G. Stelzer, A. Harel, D. Lancet, GeneCards Version 3: the human gene integrator, Database 2010 (2010) (baq020).
[24] W.R. Pearson, D.J. Lipman, Improved tools for biological sequence comparison, Proc. Natl. Acad. Sci. USA 85 (1988) 2444–2448.
[25] E. Demir, M.P. Cary, S. Paley, K. Fukuda, C. Lemer, I. Vastrik, G. Wu, P. D'Eustachio, C. Schaefer, J. Luciano, F. Schacherer, I. Martinez-Flores, Z. Hu, V. Jimenez-Jacinto, G. Joshi-Tope, K. Kandasamy, A.C. Lopez-Fuentes, H. Mi, E. Pichler, I. Rodchenkov, A. Splendiani, S. Tkachev, J. Zucker, G. Gopinath, H. Rajasimha, R. Ramakrishnan, I. Shah, M. Syed, N. Anwar, O. Babur, M. Blinov, E. Brauner, D. Corwin, S. Donaldson, F. Gibbons, R. Goldberg, P. Hornbeck, A. Luna, P. Murray-Rust, E. Neumann, O. Reubenacker, M. Samwald, M. van Iersel, S. Wimalaratne, K. Allen, B. Braun, M. Whirl-Carrillo, K.H. Cheung, K. Dahlquist, A. Finney, M. Gillespie, E. Glass, L. Gong, R. Haw, M. Honig, O. Hubaut, D. Kane, S. Krupa, M. Kutmon, J. Leonard, D. Marks, D. Merberg, V. Petri, A. Pico, D. Ravenscroft, L. Ren, N. Shah, M. Sunshine, R. Tang, R. Whaley, S. Letovksy, K.H. Buetow, A. Rzhetsky, V. Schachter, B.S. Sobral, U. Dogrusoz, S. McWeeney, M. Aladjem, E. Birney, J. Collado-Vides, S. Goto, M. Hucka, N. Le Novere, N. Maltsev, A. Pandey, P. Thomas, E. Wingender, P.D. Karp, C. Sander, G.D. Bader, The BioPAX community standard for pathway data sharing, Nat. Biotechnol. 28 (2010) 935–942.
[26] M. Kanehisa, S. Goto, Y. Sato, M. Furumichi, M. Tanabe, KEGG for integration and interpretation of large-scale molecular data sets, Nucleic Acids Res. 40 (2012) D109–D114.
[27] M. Hucka, A. Finney, H.M. Sauro, H. Bolouri, J.C. Doyle, H. Kitano, A.P. Arkin, B.J. Bornstein, D. Bray, A. Cornish-Bowden, A.A. Cuellar, S. Dronov, E.D. Gilles, M. Ginkel, V. Gor Goryanin II, W.J. Hedley, T.C. Hodgman, J.H. Hofmeyr, P.J. Hunter, N.S. Juty, J.L. Kasberger, A. Kremling, U. Kummer, N. Le Novere, L.M. Loew, D. Lucio, P. Mendes, E. Minch, E.D. Mjolsness, Y. Nakayama, M.R. Nelson, P.F. Nielsen, T. Sakurada, J.C. Schaff, B.E. Shapiro, T.S. Shimizu, H.D. Spence, J. Stelling, K. Takahashi, M. Tomita, J. Wagner, J. Wang, The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models, Bioinformatics 19 (2003) 524–531.
[28] K. Van Roey, S. Orchard, S. Kerrien, M. Dumousseau, S. Ricard-Blum, H. Hermjakob, T.J. Gibson, Capturing cooperative interactions with the PSI-MI format, Database 2013 (2013) (bat066).
[29] W. Huang da, B.T. Sherman, R.A. Lempicki, Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources, Nat. Protoc. 4 (2009) 44–57.
[30] P. Tieri, A. de la Fuente, A. Termanini, C. Franceschi, Integrating omics data for signaling pathways, interactome reconstruction, and functional analysis, Methods Mol. Biol. 719 (2011) 415–433.
[31] N. Gehlenborg, S.I. O'Donoghue, N.S. Baliga, A. Goesmann, M.A. Hibbs, H. Kitano, O. Kohlbacher, H. Neuweger, R. Schneider, D. Tenenbaum, A.C. Gavin, Visualization of omics data for systems biology, Nat. Methods 7 (2010) S56–S68.
[32] C. Pastrello, D. Otasek, K. Fortney, G. Agapito, M. Cannataro, E. Shirdel, I. Jurisica, Visual data mining of biological networks: one size does not fit all, PLoS Comput. Biol. 9 (2013) e1002833.
[33] K. Wang, I. Lee, G. Carlson, L. Hood, D. Galas, Systems biology and the discovery of diagnostic biomarkers, Dis. Markers 28 (2010) 199–207.
[34] M. Schena, D. Shalon, R.W. Davis, P.O. Brown, Quantitative monitoring of gene expression patterns with a complementary DNA microarray, Science 270 (1995) 467–470.
[35] T.R. Golub, D.K. Slonim, P. Tamayo, C. Huard, M. Gaasenbeek, J.P. Mesirov, H. Coller, M.L. Loh, J.R. Downing, M.A. Caligiuri, C.D. Bloomfield, E.S. Lander, Molecular classification of cancer: class discovery and class prediction by gene expression monitoring, Science 286 (1999) 531–537.
[36] Y. Wang, J.G. Klijn, Y. Zhang, A.M. Sieuwerts, M.P. Look, F. Yang, D. Talantov, M. Timmermans, M.E. Meijer-van Gelder, J. Yu, T. Jatkoe, E.M. Berns, D. Atkins, J.A. Foekens, Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer, Lancet 365 (2005) 671–679.
[37] G. Danaei, S. Vander Hoorn, A.D. Lopez, C.J. Murray, M. Ezzati, Causes of cancer in the world: comparative risk assessment of nine behavioural and environmental risk factors, Lancet 366 (2005) 1784–1793.
[38] A. Jemal, F. Bray, M.M. Center, J. Ferlay, E. Ward, D. Forman, Global cancer statistics, CA Cancer J. Clin. 61 (2011) 69–90.
[39] D.M. Parkin, International variation, Oncogene 23 (2004) 6329–6340.
[40] D. Zaridze, E. Borisova, D. Maximovitch, V. Chkhikvadze, Alcohol consumption, smoking and risk of gastric cancer: case–control study from Moscow, Russia, Cancer Causes Control 11 (2000) 363–371.
[41] B. Park, A. Shin, S.K. Park, K.P. Ko, S.H. Ma, E.H. Lee, J. Gwack, E.J. Jung, L.Y. Cho, J.J. Yang, K.Y. Yoo, Ecological study for refrigerator use, salt, vegetable, and fruit intakes, and gastric cancer, Cancer Causes Control 22 (2011) 1497–1502.
[42] C.A. Gonzalez, G. Pera, A. Agudo, D. Palli, V. Krogh, P. Vineis, R. Tumino, S. Panico, G. Berglund, H. Siman, O. Nyren, A. Agren, C. Martinez, M. Dorronsoro, A. Barricarte, M.J. Tormo, J.R. Quiros, N. Allen, S. Bingham, N. Day, A. Miller, G. Nagel, H. Boeing, K. Overvad, A. Tjonneland, H.B. Bueno-De-Mesquita, H.C. Boshuizen, P. Peeters, M. Numans, F. Clavel-Chapelon, I. Helen, E. Agapitos, E. Lund, M. Fahey, R. Saracci, R. Kaaks, E. Riboli, Smoking and the risk of gastric cancer in the european prospective investigation into cancer and nutrition (EPIC), Int. J. Cancer 107 (2003) 629–634.
[43] S. Bevan, R.S. Houlston, Genetic predisposition to gastric cancer, QJM 92 (1999) 5–10.
[44] M. Santibanez, J. Alguacil, M.G. de la Hera, E.M. Navarrete-Munoz, J. Llorca, N. Aragones, T. Kauppinen, J. Vioque, Occupational exposures and risk of stomach cancer by histological type, Occup. Environ. Med. 69 (2012) 268–275.
[45] Schistosomes, liver flukes and *Helicobacter pylori*. IARC working group on the evaluation of carcinogenic risks to humans. Lyon, 7–14 June 1994, IARC Monographs On The Evaluation Of Carcinogenic Risks To Humans/World Health Organization, International Agency for Research on Cancer 61 (1994) 1–241.
[46] S. Nagini, Carcinoma of the stomach: a review of epidemiology, pathogenesis, molecular genetics and chemoprevention, World J. Gastrointest. Oncol. 4 (2012) 156–169.
[47] L.M. Coussens, Z. Werb, Inflammation and cancer, Nature 420 (2002) 860–867.
[48] X. Xu, Z. Liu, M. Fang, H. Yu, X. Liang, X. Li, X. Liu, C. Chen, J. Jia, *Helicobacter pylori* CagA induces ornithine decarboxylase upregulation via Src/MEK/ERK/c-Myc pathway: implication for progression of gastric diseases, Exp. Biol. Med. 237 (2012) 435–441.
[49] G. Nardone, S. Staibano, A. Rocco, E. Mezza, F.P. D'Armiento, L. Insabato, A. Coppola, G. Salvatore, A. Lucariello, N. Figura, G. De Rosa, G. Budillon, Effect of *Helicobacter pylori* infection and its eradication on cell proliferation, DNA status, and oncogene expression in patients with chronic gastritis, Gut 44 (1999) 789–799.
[50] S.Z. Ding, J.B. Goldberg, M. Hatakeyama, *Helicobacter pylori* infection, oncogenic pathways and epigenetic mechanisms in gastric carcinogenesis, Future Oncol. 6 (2010) 851–862.
[51] T. Ushijima, N. Hattori, Molecular pathways: involvement of *Helicobacter pylori*-triggered inflammation in the formation of an epigenetic field defect, and its usefulness as cancer risk and exposure markers, Clin. Cancer Res. 18 (2012) 923–929.
[52] K. Matsushima, H. Isomoto, N. Inoue, T. Nakayama, T. Hayashi, M. Nakayama, K. Nakao, T. Hirayama, S. Kohno, MicroRNA signatures in *Helicobacter pylori*-infected gastric mucosa, Int. J. Cancer 128 (2011) 361–370.

[53] R.M. O'Connell, D.S. Rao, D. Baltimore, MicroRNA regulation of inflammatory responses, Annu. Rev. Immunol. 30 (2012) 295–312.
[54] M.V. Iorio, C.M. Croce, MicroRNA involvement in human cancer, Carcinogenesis 33 (2012) 1126–1133.
[55] J. Zabaleta, M.C. Camargo, M.B. Piazuelo, E. Fontham, B.G. Schneider, L.A. Sicinschi, W. Ferrante, L. Balart, P. Correa, A.C. Ochoa, Association of interleukin-1beta gene polymorphisms with precancerous gastric lesions in African Americans and Caucasians, Am. J. Gastroenterol. 101 (2006) 163–171.
[56] J. Zabaleta, M.C. Camargo, M.D. Ritchie, M.B. Piazuelo, R.A. Sierra, S.D. Turner, A. Delgado, E.T. Fontham, B.G. Schneider, P. Correa, A.C. Ochoa, Association of haplotypes of inflammation-related genes with gastric preneoplastic lesions in African Americans and Caucasians, Int. J. Cancer 128 (2011) 668–675.
[57] Y. Shi, Y. Zhou, The role of surgery in the treatment of gastric cancer, J. Surg. Oncol. 101 (2010) 687–692.
[58] P. Malfertheiner, F. Megraud, C.A. O'Morain, J. Atherton, A.T. Axon, F. Bazzoli, G.F. Gensini, J.P. Gisbert, D.Y. Graham, T. Rokkas, E.M. El-Omar, E.J. Kuipers, Management of *Helicobacter pylori* infection – the Maastricht IV/Florence Consensus Report, Gut 61 (2012) 646–664.
[59] P. Malfertheiner, Author's response: *Helicobacter pylori* eradication and gastric cancer prevention, Gut 62 (2013) 950–951.
[60] Z. Wang, Y. Yu, W. Yang, B. Chen, X. Li, Does *Helicobacter pylori* eradication really reduce the risk of gastric cancer at the population level?, Gut 62 (2013) 950
[61] A. Mueller, D.S. Merrell, J. Grimm, S. Falkow, Profiling of microdissected gastric epithelial cells reveals a cell type-specific response to *Helicobacter pylori* infection, Gastroenterology 127 (2004) 1446–1462.
[62] Y. Qu, S. Dang, P. Hou, Gene methylation in gastric cancer, Clin. Chim. Acta 424 (2013) 53–65.
[63] Y. Tsukamoto, T. Uchida, S. Karnan, T. Noguchi, L.T. Nguyen, M. Tanigawa, I. Takeuchi, K. Matsuura, N. Hijiya, C. Nakada, T. Kishida, K. Kawahara, H. Ito, K. Murakami, T. Fujioka, M. Seto, M. Moriyama, Genome-wide analysis of DNA copy number alterations and gene expression in gastric cancer, J. Pathol. 216 (2008) 471–482.
[64] J.G. Sinkovics, Molecular biology of oncogenic inflammatory processes. I. Non-oncogenic and oncogenic pathogens, intrinsic inflammatory reactions without pathogens, and microRNA/DNA interactions (Review), Int. J. Oncol. 40 (2012) 305–349.
[65] J. Zabaleta, Multifactorial etiology of gastric cancer, Methods Mol. Biol. 863 (2012) 411–435.
[66] Y. Wu, H. Grabsch, T. Ivanova, I.B. Tan, J. Murray, C.H. Ooi, A.I. Wright, N.P. West, G.G. Hutchins, J. Wu, M. Lee, J. Lee, J.H. Koo, K.G. Yeoh, N. van Grieken, B. Ylstra, S.Y. Rha, J.A. Ajani, J.H. Cheong, S.H. Noh, K.H. Lim, A. Boussioutas, J.S. Lee, P. Tan, Comprehensive genomic meta-analysis identifies intra-tumoural stroma as a predictor of survival in patients with gastric cancer, Gut 62 (2013) 1100–1111.
[67] S.D. Georgopoulos, V. Papastergiou, S. Karatapanis, Current options for the treatment of *Helicobacter pylori*, Expert Opin. Pharmacother. 14 (2013) 211–223.
[68] K.R. Brown, D. Otasek, M. Ali, M.J. McGuffin, W. Xie, B. Devani, I.L. Toch, I. Jurisica, NAViGaTOR: network analysis, visualization and graphing Toronto, Bioinformatics 25 (2009) 3327–3329.
[69] A. Djebbari, M. Ali, D. Otasek, M. Kotlyar, K. Fortney, S. Wong, A. Hrvojic, I. Jurisica, NAViGaTOR: scalable and interactive navigation and analysis of large graphs, Internet Math. 7 (2011) 314–347.
[70] E.A. Shirdel, W. Xie, T.W. Mak, I. Jurisica, NAViGaTing the micronome – using multiple microRNA prediction databases to identify signalling pathway-associated microRNAs, PLoS One 6 (2011) e17429.
[71] Y.J. Bang, E. Van Cutsem, A. Feyereislova, H.C. Chung, L. Shen, A. Sawaki, F. Lordick, A. Ohtsu, Y. Omuro, T. Satoh, G. Aprile, E. Kulikov, J. Hill, M. Lehle, J. Ruschoff, Y.K. Kang, Trastuzumab in combination with chemotherapy versus chemotherapy alone for treatment of HER2-positive advanced gastric or gastro-oesophageal junction cancer (ToGA): a phase 3, open-label, randomised controlled trial, Lancet 376 (2010) 687–697.
[72] A. Gamboa-Dominguez, C. Dominguez-Fonseca, L. Quintanilla-Martinez, E. Reyes-Gutierrez, D. Green, A. Angeles-Angeles, R. Busch, C. Hermannstadter, J. Nahrig, K.F. Becker, I. Becker, H. Hofler, F. Fend, B. Luber, Epidermal growth factor receptor expression correlates with poor survival in gastric adenocarcinoma from Mexican patients: a multivariate analysis using a standardized immunohistochemical detection system, Mod. Pathol. 17 (2004) 579–587.
[73] A. Atmaca, D. Werner, C. Pauligk, K. Steinmetz, R. Wirtz, H.M. Altmannsberger, E. Jager, S.E. Al-Batran, The prognostic impact of epidermal growth factor receptor in patients with metastatic gastric cancer, BMC Cancer 12 (2012) 524.
[74] L. Zhang, J. Yang, J. Cai, X. Song, J. Deng, X. Huang, D. Chen, M. Yang, J.P. Wery, S. Li, A. Wu, Z. Li, Y. Liu, Y. Chen, Q. Li, J. Ji, A subset of gastric cancers with EGFR amplification and overexpression respond to cetuximab therapy, Sci. Rep. 3 (2013) 2992.
[75] F. Sclafani, A. Roy, D. Cunningham, A. Wotherspoon, C. Peckitt, D. Gonzalez de Castro, J. Tabernero, B. Glimelius, A. Cervantes, Z. Eltahir, J. Oates, I. Chau, HER2 in high-risk rectal cancer patients treated in EXPERT-C, a randomized phase II trial of neoadjuvant capecitabine and oxaliplatin (CAPOX) and chemoradiotherapy (CRT) with or without cetuximab, Ann. Oncol. 24 (2013) 3123–3128.
[76] E. Rimbara, L.A. Fischbach, D.Y. Graham, Optimal therapy for *Helicobacter pylori* infections, Nat. Rev. Gastroenterol. Hepatol. 8 (2011) 79–88.
[77] G.H. Kang, H.J. Lee, K.S. Hwang, S. Lee, J.H. Kim, J.S. Kim, Aberrant CpG island hypermethylation of chronic gastritis, in relation to aging, gender, intestinal metaplasia, and chronic inflammation, Am. J. Pathol. 163 (2003) 1551–1556.
[78] L.A. Fischbach, P. Correa, H. Ramirez, J.L. Realpe, T. Collazos, B. Ruiz, L.E. Bravo, J.C. Bravo, A.L. Casabon, B.A. Schmidt, Anti-inflammatory and tissue-protectant drug effects: results from a randomized placebo-controlled trial of gastritis patients at high risk for gastric cancer, Aliment. Pharmacol. Ther. 15 (2001) 831–841.
[79] P.V. Missiuro, K. Liu, L. Zou, B.C. Ross, G. Zhao, J.S. Liu, H. Ge, Information flow analysis of interactome networks, PLoS Comput. Biol. 5 (2009) e1000350.
[80] S.H. Yeh, H.Y. Yeh, V.W. Soo, A network flow approach to predict drug targets from microarray data, disease genes and interactome network – case study on prostate cancer, J. Clin. Bioinformatics 2 (2012) 1.
[81] A. Pavlogiannis, V. Mozhayskiy, I. Tagkopoulos, A flood-based information flow analysis and network minimization method for gene regulatory networks, BMC Bioinformatics 14 (2013) 137.
[82] W. Lu, K. Pan, L. Zhang, D. Lin, X. Miao, W. You, Genetic polymorphisms of interleukin (IL)-1B, IL-1RN, IL-8, IL-10 and tumor necrosis factor alpha and risk of gastric cancer in a Chinese population, Carcinogenesis 26 (2005) 631–636.
[83] T. Nishizawa, H. Suzuki, The role of microRNA in gastric malignancy, Int. J. Mol. Sci. 14 (2013) 9487–9496.
[84] J. Cazalis, C. Bodet, G. Gagnon, D. Grenier, Doxycycline reduces lipopolysaccharide-induced inflammatory mediator secretion in macrophage and ex vivo human whole blood models, J. Periodontol. 79 (2008) 1762–1768.
[85] H. Jeong, S.P. Mason, A.L. Barabasi, Z.N. Oltvai, Lethality and centrality in protein networks, Nature 411 (2001) 41–42.
[86] N. Przulj, D.A. Wigle, I. Jurisica, Functional topology in a network of protein interactions, Bioinformatics 20 (2004) 340–348.
[87] T. Milenkoviae, N. Przulj, Uncovering biological network function via graphlet degree signatures, Cancer Info. 6 (2008) 257–273.
[88] P.F. Jonsson, P.A. Bates, Global topological features of cancer proteins in the human interactome, Bioinformatics 22 (2006) 2291–2297.
[89] M. Kotlyar, K. Fortney, I. Jurisica, Network-based characterization of drug-regulated genes, drug targets, and toxicity, Methods 57 (2012) 499–507.
[90] R.S. Ismail, R.L. Baldwin, J. Fang, D. Browning, B.Y. Karlan, J.C. Gasson, D.D. Chang, Differential gene expression between normal and tumor-derived ovarian epithelial cells, Cancer Res. 60 (2000) 6744–6749.
[91] K.W. Qin, H. Shi, L. Zhang, P.F. Liu, W.L. Cai, K.H. Wu, X.C. Zhu, F.F. Lan, The research on screening differentially expressed genes in Hirschsprung's disease by using microarray, J. Pediatr. Surg. 48 (2013) 2281–2288.
[92] E. Kettunen, S. Anttila, J.K. Seppanen, A. Karjalainen, H. Edgren, I. Lindstrom, R. Salovaara, A.M. Nissen, J. Salo, K. Mattson, J. Hollmen, S. Knuutila, H. Wikman, Differentially expressed genes in nonsmall cell lung cancer: expression profiling of cancer-related genes in squamous cell lung cancer, Cancer Genet. Cytogenet. 149 (2004) 98–106.
[93] K. Fortney, I. Jurisica, Integrative computational biology for cancer research, Human Genet. (2011).
[94] M. Zhang, C. Yao, Z. Guo, J. Zou, L. Zhang, H. Xiao, D. Wang, D. Yang, X. Gong, J. Zhu, Y. Li, X. Li, Apparently low reproducibility of true differential expression discoveries in microarray studies, Bioinformatics 24 (2008) 2057–2063.
[95] P.C. Boutros, S.K. Lau, M. Pintilie, N. Liu, F.A. Shepherd, S.D. Der, M.S. Tsao, L.Z. Penn, I. Jurisica, Prognostic gene signatures for non-small-cell lung cancer, Proc. Natl. Acad. Sci. USA 106 (2009) 2824–2828.
[96] H.Y. Chuang, E. Lee, Y.T. Liu, D. Lee, T. Ideker, Network-based classification of breast cancer metastasis, Mol. Syst. Biol. 3 (2007) 140.
[97] K. Fortney, M. Kotlyar, I. Jurisica, Inferring the functions of longevity genes with modular subnetwork biomarkers of *Caenorhabditis elegans* aging, Genome Biol. 11 (2010) R13.
[98] T. Ideker, O. Ozier, B. Schwikowski, A.F. Siegel, Discovering regulatory and signalling circuits in molecular interaction networks, Bioinformatics 18 (Suppl. 1) (2002) S233–S240.
[99] N. Przulj, Biological network comparison using graphlet degree distribution, Bioinformatics 23 (2007) e177–e183.
[100] D.B. West, Introduction to Graph Theory, second ed., Prentice Hall, Upper Saddle River, NJ, 2001.
[101] N. Przulj, Analyzing Large Biological Networks: Protein–Protein Interactions Example, Department of Computer Science, University of Toronto, Toronto, 2005.
[102] D.J. Watts, S.H. Strogatz, Collective dynamics of 'small-world' networks, Nature 393 (1998) 440–442.
[103] O. Mason, M. Verwoerd, Graph theory and networks in biology, IET Syst. Biol. 1 (2007) 89–119.
[104] L.C. Freeman, A set of measures of centrality based on betweenness, Sociometry 40 (1977) 35–41.
[105] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, U. Alon, Network motifs: simple building blocks of complex networks, Science 298 (2002) 824–827.
[106] N. Przulj, D.G. Corneil, I. Jurisica, Modeling interactome: scale-free or geometric?, Bioinformatics 20 (2004) 3508–3515
[107] M. Garey, D.S. Johnson, Computers and Intractability: A Guide to the Theory of NP-Completeness, W.H. Freeman and Company, 1979.
[108] D. Marcus, Y. Shavitt, RAGE: a rapid graphlet enumerator for large networks, Comput. Netw. 56 (2012) 810–819.
[109] N. Shervashidze, S.V.N. Viswanathan, T. Petri, K. Mehlhorn, K.M. Borgwardt, Efficient graphlet kernels for large graph comparison, in: Proceedings of the

12th International Conference on Artificial Intelligence and Statistics, 2009, pp. 488–495.
[110] N. Kashtan, S. Itzkovitz, R. Milo, U. Alon, Efficient sampling algorithm for estimating subgraph concentrations and detecting network motifs, Bioinformatics 20 (2004) 1746–1758.
[111] S. Wernicke, Efficient detection of network motifs, IEEE/ACM Trans. Comput. Biol. Bioinform. 3 (2006) 347–359.
[112] N. Przulj, D.G. Corneil, I. Jurisica, Efficient estimation of graphlet frequency distributions in protein–protein interaction networks, Bioinformatics 22 (2006) 974–980.
[113] S. Mostafavi, D. Ray, D. Warde-Farley, C. Grouios, Q. Morris, GeneMANIA: a real-time multiple association network integration algorithm for predicting gene function, Genome Biol. 9 (Suppl. 1) (2008) S4.
[114] O. Vanunu, O. Magger, E. Ruppin, T. Shlomi, R. Sharan, Associating genes and protein complexes with disease via network propagation, PLoS Comput. Biol. 6 (2010) e1000641.
[115] A.L. Barabasi, N. Gulbahce, J. Loscalzo, Network medicine: a network-based approach to human disease, Nat. Rev. Genet. 12 (2011) 56–68.
[116] A.M. Edwards, R. Isserlin, G.D. Bader, S.V. Frye, T.M. Willson, F.H. Yu, Too many roads not taken, Nature 470 (2011) 163–165.
[117] C. von Mering, R. Krause, B. Snel, M. Cornell, S.G. Oliver, S. Fields, P. Bork, Comparative assessment of large-scale data sets of protein–protein interactions, Nature 417 (2002) 399–403.
[118] S.J. Wodak, S. Pu, J. Vlasblom, B. Seraphin, Challenges and rewards of interaction proteomics, Mol. Cell. Proteomics (2008).
[119] L. Hakes, J.W. Pinney, D.L. Robertson, S.C. Lovell, Protein–protein interaction networks and biology – what's the connection?, Nat Biotechnol. 26 (2008) 69–72.
[120] R. Albert, H. Jeong, A.L. Barabasi, Error and attack tolerance of complex networks, Nature 406 (2000) 378–382.
[121] L. Hakes, D.L. Robertson, S.G. Oliver, Effect of dataset selection on the topological interpretation of protein interaction networks, BMC Genomics 6 (2005) 131.
[122] R. Mrowka, A. Patzak, H. Herzel, Is there a bias in proteome research?, Genome Res 11 (2001) 1971–1973.
[123] E. Sprinzak, S. Sattath, H. Margalit, How reliable are experimental protein–protein interaction data?, J Mol. Biol. 327 (2003) 919–923.
[124] K.R. Brown, I. Jurisica, Unequal evolutionary conservation of human protein interactions in interologous networks, Genome Biol. 8 (2007) R95.
[125] O. Kuchaiev, A. Stevanovic, W. Hayes, N. Przulj, GraphCrunch 2: software tool for network modeling, alignment and clustering, BMC Bioinformatics 12 (2011) 24.
[126] B.A. Kelch, D.L. Makino, M. O'Donnell, J. Kuriyan, How a DNA polymerase clamp loader opens a sliding clamp, Science 334 (2011) 1675–1680.
[127] D.S. Leggett, J. Hanna, A. Borodovsky, B. Crosas, M. Schmidt, R.T. Baker, T. Walz, H. Ploegh, D. Finley, Multiple associated proteins regulate proteasome structure and function, Mol. Cell 10 (2002) 495–507.
[128] M. Fetchko, I. Stagljar, Application of the split-ubiquitin membrane yeast two-hybrid system to investigate membrane protein interactions, Methods 32 (2004) 349–362.
[129] M. Barrios-Rodiles, K.R. Brown, B. Ozdamar, R. Bose, Z. Liu, R.S. Donovan, F. Shinjo, Y. Liu, J. Dembowy, I.W. Taylor, V. Luga, N. Przulj, M. Robinson, H. Suzuki, Y. Hayashizaki, I. Jurisica, J.L. Wrana, High-throughput mapping of a dynamic signaling network in mammalian cells, Science 307 (2005) 1621–1625.
[130] J. Rachlin, D.D. Cohen, C. Cantor, S. Kasif, Biological context networks: a mosaic view of the interactome, Mol. Syst. Biol. 2 (2006) 66.
[131] M.H. Schaefer, T.J. Lopes, N. Mah, J.E. Shoemaker, Y. Matsuoka, J.F. Fontaine, C. Louis-Jeune, A.J. Eisfeld, G. Neumann, C. Perez-Iratxeta, Y. Kawaoka, H. Kitano, M.A. Andrade-Navarro, Adding protein context to the human protein–protein interaction network to reveal meaningful interactions, PLoS Comput. Biol. 9 (2013) e1002860.
[132] B. Lehner, C.M. Sanderson, A protein interaction framework for human mRNA degradation, Genome Res. 14 (2004) 1315–1323.
[133] C. Behrends, M.E. Sowa, S.P. Gygi, J.W. Harper, Network organization of the human autophagy system, Nature 466 (2010) 68–76.
[134] A. Hegele, A. Kamburov, A. Grossmann, C. Sourlis, S. Wowro, M. Weimann, C.L. Will, V. Pena, R. Luhrmann, U. Stelzl, Dynamic protein–protein interaction wiring of the human spliceosome, Mol. Cell 45 (2012) 567–580.
[135] J. Li, K. Bennett, A. Stukalov, B. Fang, G. Zhang, T. Yoshida, I. Okamoto, J.Y. Kim, L. Song, Y. Bai, X. Qian, B. Rawal, M. Schell, F. Grebien, G. Winter, U. Rix, S. Eschrich, J. Colinge, J. Koomen, G. Superti-Furga, E.B. Haura, Perturbation of the mutated EGFR interactome identifies vulnerabilities and resistance mechanisms, Mol. Syst. Biol. 9 (2013) 705.
[136] A. Bossi, B. Lehner, Tissue specificity and the human protein interaction network, Mol. Syst. Biol. 5 (2009) 260.
[137] D. Emig, M. Albrecht, Tissue-specific proteins and functional implications, J. Proteome Res. 10 (2011) 1893–1903.
[138] J.D. Han, N. Bertin, T. Hao, D.S. Goldberg, G.F. Berriz, L.V. Zhang, D. Dupuy, A.J. Walhout, M.E. Cusick, F.P. Roth, M. Vidal, Evidence for dynamically organized modularity in the yeast protein–protein interaction network, Nature 430 (2004) 88–93.
[139] W.H. Lin, W.C. Liu, M.J. Hwang, Topological and organizational properties of the products of house-keeping and tissue-specific genes in protein–protein interaction networks, BMC Syst. Biol. 3 (2009) 32.
[140] S. Agarwal, C.M. Deane, M.A. Porter, N.S. Jones, Revisiting date and party hubs: novel approaches to role assignment in protein interaction networks, PLoS Comput. Biol. 6 (2010) e1000817.
[141] S.J. Xiao, C. Zhang, Q. Zou, Z.L. Ji, TiSGeD: a database for tissue-specific genes, Bioinformatics 26 (2010) 1273–1275.
[142] R. Barshir, O. Basha, A. Eluk, I.Y. Smoly, A. Lan, E. Yeger-Lotem, The TissueNet database of human tissue protein–protein interactions, Nucleic Acids Res. 41 (2013) D841–D844.
[143] A.I. Su, T. Wiltshire, S. Batalov, H. Lapp, K.A. Ching, D. Block, J. Zhang, R. Soden, M. Hayakawa, G. Kreiman, M.P. Cooke, J.R. Walker, J.B. Hogenesch, A gene atlas of the mouse and human protein-encoding transcriptomes, Proc. Natl. Acad. Sci. USA 101 (2004) 6062–6067.
[144] C. Conrad, D.W. Gerlich, Automated microscopy for high-content RNAi screening, J. Cell Biol. 188 (2010) 453–461.
[145] L.H. Loo, L.F. Wu, S.J. Altschuler, Image-based multivariate profiling of drug responses from single cells, Nat. Methods 4 (2007) 445–453.
[146] B.L. Millard, M. Niepel, M.P. Menden, J.L. Muhlich, P.K. Sorger, Adaptive informatics for multifactorial and high-content biological data, Nat. Methods 8 (2011) 487–493.
[147] K.A. Janes, M.B. Yaffe, Data-driven modelling of signal-transduction networks, Nat. Rev. Mol. Cell Biol. 7 (2006) 820–828.
[148] K.A. Janes, J.G. Albeck, S. Gaudet, P.K. Sorger, D.A. Lauffenburger, M.B. Yaffe, A systems model of signaling identifies a molecular basis set for cytokine-induced apoptosis, Science 310 (2005) 1646–1653.
[149] N. Kumar, R. Afeyan, H.D. Kim, D.A. Lauffenburger, Multipathway model enables prediction of kinase inhibitor cross-talk effects on migration of Her2-overexpressing mammary epithelial cells, Mol. Pharmacol. 73 (2008) 1668–1678.
[150] C. Jorgensen, A. Sherman, G.I. Chen, A. Pasculescu, A. Poliakov, M. Hsiung, B. Larsen, D.G. Wilkinson, R. Linding, T. Pawson, Cell-specific information processing in segregating populations of Eph receptor ephrin-expressing cells, Science 326 (2009) 1502–1509.
[151] L.G. Alexopoulos, J. Saez-Rodriguez, B.D. Cosgrove, D.A. Lauffenburger, P.K. Sorger, Networks inferred from biochemical data reveal profound differences in toll-like receptor and inflammatory signaling between normal and transformed hepatocytes, Mol. Cell. Proteomics 9 (2010) 1849–1865.
[152] B.B. Aldridge, J.M. Burke, D.A. Lauffenburger, P.K. Sorger, Physicochemical modelling of cell signalling pathways, Nat. Cell Biol. 8 (2006) 1195–1203.
[153] U.S. Bhalla, R. Iyengar, Emergent properties of networks of biological signaling pathways, Science 283 (1999) 381–387.
[154] B.N. Kholodenko, O.V. Demin, G. Moehren, J.B. Hoek, Quantification of short term signaling by the epidermal growth factor receptor, J. Biol. Chem. 274 (1999) 30169–30181.
[155] B. Schoeberl, C. Eichler-Jonsson, E.D. Gilles, G. Muller, Computational modeling of the dynamics of the MAP kinase cascade activated by surface and internalized EGF receptors, Nat. Biotechnol. 20 (2002) 370–375.
[156] N. Borisov, E. Aksamitiene, A. Kiyatkin, S. Legewie, J. Berkhout, T. Maiwald, N.P. Kaimachnikov, J. Timmer, J.B. Hoek, B.N. Kholodenko, Systems-level interactions between insulin–EGF networks amplify mitogenic signaling, Mol. Syst. Biol. 5 (2009) 256.
[157] M.R. Birtwistle, M. Hatakeyama, N. Yumoto, B.A. Ogunnaike, J.B. Hoek, B.N. Kholodenko, Ligand-dependent responses of the ErbB signaling network: experimental and modeling analyses, Mol. Syst. Biol. 3 (2007) 144.
[158] M. Hatakeyama, S. Kimura, T. Naka, T. Kawasaki, N. Yumoto, M. Ichikawa, J.H. Kim, K. Saito, M. Saeki, M. Shirouzu, S. Yokoyama, A. Konagaya, A computational model on the modulation of mitogen-activated protein kinase (MAPK) and Akt pathways in heregulin-induced ErbB signalling, Biochem. J. 373 (2003) 451–463.
[159] A. Kiyatkin, E. Aksamitiene, N.I. Markevich, N.M. Borisov, J.B. Hoek, B.N. Kholodenko, Scaffolding protein Grb2-associated binder 1 sustains epidermal growth factor-induced mitogenic and survival signaling by multiple positive feedback loops, J. Biol. Chem. 281 (2006) 19925–19938.
[160] B.S. Hendriks, L.K. Opresko, H.S. Wiley, D. Lauffenburger, Quantitative analysis of HER2-mediated effects on HER2 and epidermal growth factor receptor endocytosis: distribution of homo- and heterodimers depends on relative HER2 levels, J. Biol. Chem. 278 (2003) 23343–23351.
[161] W.W. Chen, B. Schoeberl, P.J. Jasper, M. Niepel, U.B. Nielsen, D.A. Lauffenburger, P.K. Sorger, Input–output behavior of ErbB signaling pathways as revealed by a mass action model trained against dynamic data, Mol. Syst. Biol. 5 (2009) 239.
[162] T. Haberichter, B. Madge, R.A. Christopher, N. Yoshioka, A. Dhiman, R. Miller, R. Gendelman, S.V. Aksenov, I.G. Khalil, S.F. Dowdy, A systems biology dynamical model of mammalian G1 cell cycle progression, Mol. Syst. Biol. 3 (2007) 84.
[163] S.Y. Shin, O. Rath, S.M. Choo, F. Fee, B. McFerran, W. Kolch, K.H. Cho, Positive- and negative-feedback regulations coordinate the dynamic behavior of the Ras–Raf–MEK–ERK signal transduction pathway, J. Cell Sci. 122 (2009) 425–435.
[164] D. Faratian, A. Goltsov, G. Lebedeva, A. Sorokin, S. Moodie, P. Mullen, C. Kay, I.H. Um, S. Langdon, I. Goryanin, D.J. Harrison, Systems biology reveals new strategies for personalizing cancer medicine and confirms the role of PTEN in resistance to trastuzumab, Cancer Res. 69 (2009) 6713–6720.
[165] R. Christopher, A. Dhiman, J. Fox, R. Gendelman, T. Haberitcher, D. Kagle, G. Spizz, I.G. Khalil, C. Hill, Data-driven computer simulation of human cancer cell, Ann. NY. Acad. Sci. 1020 (2004) 132–153.

[166] B. Goldstein, J.R. Faeder, W.S. Hlavacek, Mathematical and computational models of immune-receptor signalling, Nat. Rev. Immunol. 4 (2004) 445–456.
[167] M. Gstaiger, R. Aebersold, Applying mass spectrometry-based proteomics to genetics, genomics and network biology, Nat. Rev. Genet. 10 (2009) 617–627.
[168] G. MacBeath, Protein microarrays and proteomics, Nat. Genet. 32 (Suppl.) (2002) 526–532.
[169] B.N. Kholodenko, Cell-signalling dynamics in time and space, Nat. Rev. Mol. Cell Biol. 7 (2006) 165–176.
[170] M.L. Blinov, J.R. Faeder, B. Goldstein, W.S. Hlavacek, BioNetGen: software for rule-based modeling of signal transduction based on the interactions of molecular domains, Bioinformatics 20 (2004) 3289–3291.
[171] D. Faratian, R.G. Clyde, J.W. Crawford, D.J. Harrison, Systems pathology-taking molecular pathology into a new dimension, Nat. Rev. Clin. Oncol. 6 (2009) 455–464.
[172] D. Faratian, S.L. Moodie, D.J. Harrison, I. Goryanin, Dynamic computational modeling in the search for better breast cancer drug therapy, Pharmacogenomics 8 (2007) 1757–1761.
[173] H. de Jong, Modeling and simulation of genetic regulatory systems: a literature review, J. Comput. Biol. 9 (2002) 67–103.
[174] J. Fisher, T.A. Henzinger, Executable cell biology, Nat. Biotechnol. 25 (2007) 1239–1249.
[175] M. Chaves, R. Albert, E.D. Sontag, Robustness and fragility of Boolean models for genetic regulatory networks, J. Theor. Biol. 235 (2005) 431–449.
[176] S. Gupta, S.S. Bisht, R. Kukreti, S. Jain, S.K. Brahmachari, Boolean network analysis of a neurotransmitter signaling pathway, J. Theor. Biol. 244 (2007) 463–469.
[177] J. Saez-Rodriguez, L. Simeoni, J.A. Lindquist, R. Hemenway, U. Bommhardt, B. Arndt, U.U. Haus, R. Weismantel, E.D. Gilles, S. Klamt, B. Schraven, A logical model provides insights into T cell receptor signaling, PLoS Comput. Biol. 3 (2007) e163.
[178] R. Zhang, M.V. Shah, J. Yang, S.B. Nyland, X. Liu, J.K. Yun, R. Albert, T.P. Loughran Jr., Network model of survival signaling in large granular lymphocyte leukemia, Proc. Natl. Acad. Sci. USA 105 (2008) 16308–16313.
[179] J. Saez-Rodriguez, L.G. Alexopoulos, J. Epperlein, R. Samaga, D.A. Lauffenburger, S. Klamt, P.K. Sorger, Discrete logic modelling as a means to link protein signalling networks with functional analysis of mammalian signal transduction, Mol. Syst. Biol. 5 (2009) 331.
[180] R. Samaga, J. Saez-Rodriguez, L.G. Alexopoulos, P.K. Sorger, S. Klamt, The logic of EGFR/ErbB signaling: theoretical properties and analysis of high-throughput data, PLoS Comput. Biol. 5 (2009) e1000438.
[181] S.V. Aksenov, B. Church, A. Dhiman, A. Georgieva, R. Sarangapani, G. Helmlinger, I.G. Khalil, An integrated approach for inference and mechanistic modeling for advancing drug development, FEBS Lett. 579 (2005) 1878–1883.
[182] T.M. Przytycka, M. Singh, D.K. Slonim, Toward the dynamic interactome: it's about time, Brief. Bioinform. 11 (2010) 15–29.
[183] S. Orchard, S. Kerrien, P. Jones, A. Ceol, A. Chatr-Aryamontri, L. Salwinski, J. Nerothin, H. Hermjakob, Submit your interaction data the IMEx way: a step by step guide to trouble-free deposition, Proteomics 7 (Suppl. 1) (2007) 28–34.
[184] M. Ercsey-Ravasz, R.N. Lichtenwalter, N.V. Chawla, Z. Toroczkai, Range-limited centrality measures in complex networks, Phys. Rev. E Stat. Nonlin Soft Matter Phys. 85 (2012) 066103.
[185] Y. Zhang, S. De, J.R. Garner, K. Smith, S.A. Wang, K.G. Becker, Systematic analysis, comparison, and integration of disease based human genetic association data and mouse genetic phenotypic information, BMC Med. Genomics 3 (2010) 1.
[186] A.P. Davis, B.L. King, S. Mockus, C.G. Murphy, C. Saraceni-Richards, M. Rosenstein, T. Wiegers, C.J. Mattingly, The comparative toxicogenomics database: update 2011, Nucleic Acids Res. 39 (2011) D1067–D1072.