D548–D554 *Nucleic Acids Research, 2016, Vol. 44, Database issue* *Published online 13 October 2015*
doi: 10.1093/nar/gkv1048

SIGNOR: a database of causal relationships between biological entities

Livia Perfetto†, Leonardo Briganti†, Alberto Calderone†, Andrea Cerquone Perpetuini, Marta Iannuccelli, Francesca Langone, Luana Licata, Milica Marinkovic, Anna Mattioni, Theodora Pavlidou, Daniele Peluso, Lucia Lisa Petrilli, Stefano Pirrò, Daniela Posca, Elena Santonico, Alessandra Silvestri, Filomena Spada, Luisa Castagnoli and Gianni Cesareni*

Department of Biology, University of Rome Tor Vergata, Rome, Italy

Received August 13, 2015; Revised September 28, 2015; Accepted October 02, 2015

ABSTRACT

Assembly of large biochemical networks can be achieved by confronting new cell-specific experimental data with an interaction subspace constrained by prior literature evidence. The SIGnaling Network Open Resource, SIGNOR (available on line at http://signor.uniroma2.it), was developed to support such a strategy by providing a scaffold of prior experimental evidence of causal relationships between biological entities. The core of SIGNOR is a collection of approximately 12 000 manually-annotated causal relationships between over 2800 human proteins participating in signal transduction. Other entities annotated in SIGNOR are complexes, chemicals, phenotypes and stimuli. The information captured in SIGNOR can be represented as a signed directed graph illustrating the activation/inactivation relationships between signalling entities. Each entry is associated to the post-translational modifications that cause the activation/inactivation of the target proteins. More than 4900 modified residues causing a change in protein concentration or activity have been curated and linked to the modifying enzymes (about 351 human kinases and 94 phosphatases). Additional modifications such as ubiquitinations, sumoylations, acetylations and their effect on the modified target proteins are also annotated. This wealth of structured information can support experimental approaches based on multi-parametric analysis of cell systems after physiological or pathological perturbations and to assemble large logic models.

INTRODUCTION

Systems-level understanding of cell physiology is facilitated by the availability of comprehensive maps representing the interactions, or functional relationships, between biomolecules. However, experimental information about biological relationships comes in different flavours and it is difficult to capture all the relevant facts in a single data-model. As an example, one can distinguish between physical interactions supported by evidence demonstrating the formation of macromolecular complexes and functional causative, interactions where the activation of one entity modulates the concentration or activity of a second entity.

Over the past decades the development of high and low throughput approaches to identify protein complexes has motivated a number of groups to develop databases capturing this type of 'physical' interaction information (1). These initiatives have recently come to maturation when a number of protein–protein interaction (PPI) resources joined their efforts and adhered to the IMEx consortium (2) that coordinates curation and more recently aims at unifying the informatics infrastructure (3).

Causal interactions, on the other hand, are more complex to capture in a structured format but are more informative and are essential when one wants to represent the direction and sign of information flow in signal transduction. In addition to their inherent higher complexity, additional complications come from the fact that there is no unique way to model signalling interactions (4,5). As illustrated in Figure 1A for instance, the activation of the protein ERK by the kinase MEK can be represented by at least two models. Process description representations, as in the metabolic networks of the KEGG pathway database (6) and pathways in the Reactome pathway database (7), allow mechanistic descriptions, making process description maps suitable representations of chemical kinetic models. These types of descriptions lend themselves to implementation in kinetic

*To whom correspondence should be addressed. Tel: +39 06 72594315; Fax: +39 06 72594599; Email: cesareni@uniroma2.it
†These authors contributed equally to the paper as the first authors.

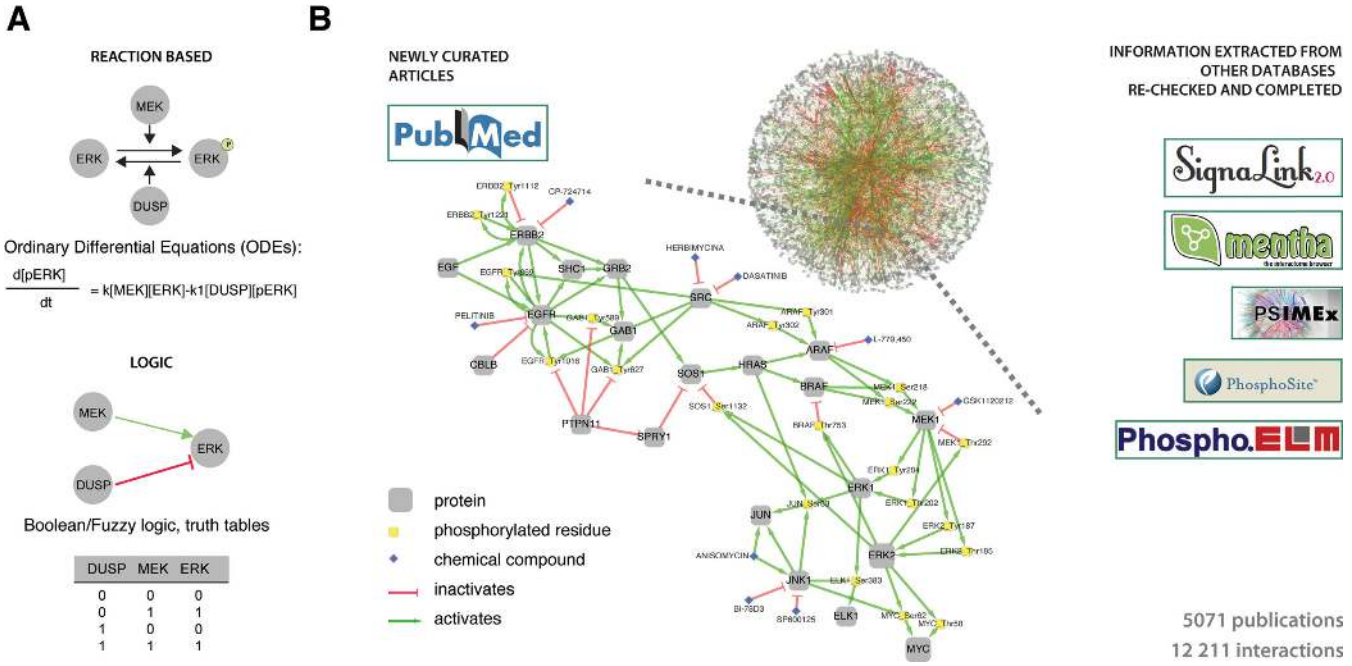

Figure 1. Logic models versus Reaction based models. (**A**) In 'reaction based models', pathways are represented as chains of chemical reactions where every variant of a component is assigned to a node, the transition of the component between two states is modulated by regulatory components. The formalism that better represents such model is constituted by ordinary differential equations (ODEs) in which each node is associated with a number that represents the concentration of the respective component. In 'logic models', nodes (molecules) are connected by directed edges, representing regulatory interactions. The state of any node depends on the state of upstream nodes and on the type of relationship that link a source to a target node. In logic networks pathways are represented as truth tables that compute the values of each node over time as a function of the states of upstream nodes. (**B**) SIGNOR stores more than 12 000 causal relationships between cellular components, originating from a de novo curation effort or from external DBs (PhosphositePlus, PhosphoELM, IMEx databases and SignaLink) (2,9,17,18). SIGNOR stores more than 4900 phosphorylation and 230 dephosphorylation reactions annotated according to the consequences (activation/inhibition) of the phosphorylation event on the target protein.

models based on systems of ordinary differential equations (ODE). The mechanistic details that can be captured by process description models, however, come to the expenses of a limitation in the size of the models that can be practically analysed.

A second type of representation is often referred to as activity flow diagrams, which are simpler and are used when the detail of a chemical reaction is not known, or is not judged essentially to understand the underlying biology and when network coverage is deemed more important than mechanistic detail. Such diagrams represent protein activities as nodes linked by directed activating or inhibitory edges without reference to the specific mechanism. The non-metabolic part of the KEGG database and the SignaLink database (8,9) use such a representation. The qualitative nature of the information is suitable for large models and, in particular, for logic models. Although such an approach yields limited mechanistic insight, the lack of parameters facilitates training to experimental data (10).

As argued later on, it is presently difficult to assemble, from the available resources, activity flow diagrams having sufficiently high coverage and linking to the experiments supporting each interaction. These considerations motivated us to develop a new database, SIGNOR, that captures causal interactions between proteins (and other biological entities) and has the same goal of accuracy, curation depth and coverage that has been achieved with protein interaction databases (11).

RESULTS

SIGNOR (http://signor.uniroma2.it), the SIGnaling Network Open Resource is a new database designed to facilitate the storage and analysis of causal interactions, i.e. interactions where a source entity has an effect (up-regulation, down-regulation, etc.) on a target entity. An ongoing curation effort in our group aims at making SIGNOR a prominent resource in the biological community by offering a comprehensive network of experimentally validated functional relationships between signalling proteins. At the time of writing, the core of SIGNOR is a collection of approximately 12 000 manually-annotated causal relationships between proteins and other biological entities that participate in signal transduction. Each relationship is linked to the literature reporting the experimental evidence and it is assigned a score. SIGNOR is not a collection of pathways but rather a collection of causal relationships. Pathways may be defined either by experts, or by users as a list of functionally related proteins. The database interrogated with a list of proteins returns graphs having as nodes the proteins in the queries list and as edges the causal relationships annotated in the database.

SIGNOR scope and data model

Cell physiology has different layers of control. Cytokines and growth factors, together with other soluble or membrane bound entities, activate cell surface receptors that in

turn trigger a signalling response that modulates the gene expression and metabolic program. The complete coverage of the metabolic and gene expression layer is beyond the current scope of SIGNOR. SIGNOR focuses on the signalling reactions between intracellular proteins and small molecules but, at the same time, it provides extensive coverage of entities, such as receptors, transcription factors and metabolic enzymes that form the interface with extracellular signalling, gene regulation and metabolism, respectively.

Causal relationships are captured in SIGNOR as a list of binary interactions between two entities, one playing the role of the regulator and the other that of the regulated entity. Most of the entities in the network are proteins. However, we also consider additional entities such as small molecules (ATP cAMP, PIP3, etc.), chemicals (enzyme inhibitors ...), stimuli, phenotypes, complexes and protein families. A protein family is a group of proteins, encoded by distinct genomic loci, with high sequence homology and sharing a redundant function (14–3–3, Erk, etc.). A complex is a group of proteins that carries out a specific function only when associated in a complex (Nfkb, mTORC1, etc). Alongside proteins and chemicals we introduce stimuli and phenotypes entities in order to be able to capture experimental evidence such as 'starvation induces the activation of AMPK' or 'mTOR inhibits autophagy'. To avoid redundancy or ambiguity between terms we provide definitions and, whenever possible, we mapped stimuli and phenotypes to Gene Ontology terms (14).

Since SIGNOR aims at being not only a repository of causal relationships but also a useful tool to conceive and design new experiments, we also annotate proteins with information about chemical inhibitors to help rational design of perturbation experiments. To this end, we manually curated the SelleckChem catalogue, representing an exhaustive repertoire of chemical compounds inhibiting protein activities. SIGNOR contains information about approximately 420 chemical inhibitors. Proteins, small molecules and chemicals are identified with UniProtKB or PubChem identifiers and hyperlinked to the respective databases (12,13). Protein complexes, protein families, stimuli and phenotypes are defined within SIGNOR. SIGNOR was conceived to be a network of human proteins. If a curated experiment supporting a causal relationship was performed in a model system, the interaction is mapped onto the human proteome maintaining the information about the experimental system (species, tissue, cell line).

Databases capturing causal relationships have not reached the maturity of the protein interaction field. Community defined and accepted standards and controlled vocabularies are not yet available. Whenever possible biological terms are mapped to Gene Ontology (14). Species are mapped to the taxid of the taxonomy database of NCBI, while cell lines and tissues are mapped to BRENDA (15). Additional control vocabulary terms are defined in Supplementary Tables S1 and S2.

Curation strategy

To populate SIGNOR we initially selected to curate manuscripts reporting signalling information by searching Medline abstracts for text containing protein names and terms typical of signalling. As a second step we filtered interactions already curated by the IMEx databases to extract some interactions that are relevant for signalling. IMEx databases do not aim to capture causal interactions. However, some of the interactions curated by these databases are relevant for signalling and imply an asymmetric relationship between the two interacting entities. These include interactions between ligands and receptors, interactions among transcription factors and enzymatic interactions. To this end we retrieved from mentha (16) the list of interactions with these characteristics and the publications supporting each interaction. A subset of these publications was classified as relevant and subjected to additional manual curation. We next checked whether phosphorylation information annotated in the PhosphoSitePlus and PhosphoELM (17,18) had already been annotated in SIGNOR. If not, we have used the literature links accompanying the missing entries to recurate the relevant information according to our curation rules. In Figure 1B we have schematized the strategy that we have adopted.

We also compared the entries in SIGNOR with those in SignaLink. We collected all the interactions that were annotated in both databases and contained contradictory information. We reviewed the literature and, if necessary, we corrected the entries in SIGNOR. Finally, the interactions that were annotated in SignaLink and absent in SIGNOR were checked for relevance according to our data model and a fraction of these were recurated according to our curation policy and curation depth. A final comparison still reveals that approximately 300 entries in SIGNOR and SignaLink provide contrasting information. We randomly selected a sample of 150 interactions and we observed that 2% needed to be revised in SIGNOR.

The curation effort described above resulted in a core of approximately 12 000 manually-annotated relationships between more than 2800 human proteins that participate in signal transduction, 321 were classified as 'receptors' (14% of the whole list of receptors), while 388 as 'transcription factors' (33% of the whole list of TFs) (19,20). Particular attention was paid to post translational modification: more than 4900 modified residues (approx. 6000 relationships) have been annotated with plenty of details about the responsible enzyme(s) and the effect of the modification on the target protein. The resulting network provides information about 351 human kinases and 94 phosphatases (21). Finally, as mentioned earlier, each relationship, in SIGNOR, is linked to the literature reporting the experimental evidence (PubMedID) and is assigned a score based on cocitation in literature abstracts. Further details regarding literature mining and data curation policy are available in the Methods session (Supplementary file 1).

SIGNOR, the website

SIGNOR can be accessed at http://signor.uniroma2.it/. The homepage summarizes the scope of the project and links to external related resources. Three different entry points provide access to SIGNOR data (i) 'entity search', (ii) 'pathway search' and (iii) 'multi-protein search'.

The 'Entity search' is the simplest way to query SIGNOR and to explore entity relationships. Proteins, inhibitors, small organic molecules or complexes can be searched by typing in the search field the entity name or its accession number (UniProtKB AC for proteins, PubChem ID for small molecules and chemicals) (12,13). SIGNOR retrieves the entities whose description matches the entered string and leaves to the user the choice of the specific entity of interest. By clicking on the selected entity the user is taken to the 'entity page' listing the information about the molecule (name aliases, accession numbers, inhibitors) and its causal relationships with other entities. If the selected entity is a protein, the user is offered a short description of the protein properties and function. This information is extracted from the UniProtKB database (12). More importantly, the 'entity page' is the entry point for browsing through the causal relationships, which are the core information of the SIGNOR database. The user can explore the relevant interactions in two ways: (i) through 'the relationships section', displaying the list of logic relationships where the searched entity plays a role as upstream regulator or as downstream effector, (ii) the graphic viewer that offers the same type of information via a graphic display (Figure 2A). In both cases a symbol and a colour code associated to the edges summarizes the attributes of the interactions. Direct interactions are displayed as solid lines, indirect as dashed lines while the effect (activation or inactivation) is mapped to the edge colour and arrow shape. Activations are represented as blue arrows, while inhibitions as red 'T-shaped' arrows. In addition, every entity pair is linked to the mechanism (phosphorylation, methylation, binding ...) and to the abstract(s) of the manuscript(s) supporting the interaction.

The 'pathway search' option allows browsing through a list of annotated pathways (see methods). In our curation effort we have defined for each pathway a set of 20–30 entities (seeds) that are often mentioned in reviews that discuss the pathway and we define the 'pathway' as the ensemble of interactions curated in SIGNOR that links the seed proteins. A drop-down menu allows the user to select one of the curated pathways recovering its graphical representation and the list of proteins that curators have annotated as 'seeds' of the pathway (Figure 2B).

The 'Multi-proteins search' function was developed to allow retrieval of relationships between a custom list of proteins or molecules. The output is the graphical representation of the network. The 'Multi-proteins search' function offers two options: the first one, 'connect', returns only the relationships between the protein seeds entered in the search field. The second option, 'all', extends the network to the first neighbours of the entities in the list. This tool allows the user to explore whether a given list of proteins are connected by activation/inactivation relationships that have been described in the literature and annotated in SIGNOR. Every search returns the results as an interactive map (Figure 2). The graphic viewer (see Supplementary File 1) provides a schematized, still detail rich, data presentation.

The top menu also gives access to database statistics and to the download page, where it is possible to download the full data set or customized data either as tab delimited text or in an SBML format. Finally two additional links permit to contact our group via email for questions, suggestions or bugs. The final link is dedicated to registered users and gives access to the curation interface.

All the entries of SIGNOR are double checked by our curators. However, we are aware that mistakes and inaccuracies could still be present. We welcome feedback from users within specific domains of their expertise. To this end a link in the home page, 'feedback', allows the users to comment on the curation of a specific entry and to suggest new entries.

CONCLUSIONS

Comparison with other signalling databases

SIGNOR (http://signor.uniroma2.it/) was developed to support experimental approaches based on multiparametric analysis of cell systems. SIGNOR offers a large network of experimentally validated causal relationships between signalling proteins and can be used as a prior model for network optimization strategies and as a basis for interpreting high content phosphoproteomic studies.

SIGNOR is different in scope from protein interaction databases. However, there is some overlap between the relationships captured by the two types of resources. Of the approximately 6500 protein interactions annotated in SIGNOR, only 45% are present in mentha that integrates about 180 000 physical interactions curated by the IMEx databases.

Because of the different data models adopted by the individual signalling databases (such as Reactome pathway database (7)) only a few of these can be fairly compared with SIGNOR. The Venn diagram in Figure 3 shows the comparison between the information captured by SIGNOR mentha, KEGG and SignaLink. Additional comparison of the characteristics of traditional signalling resources and SIGNOR is provided in Supplementary Table S3. The small overlap between the database contents is partly the result of the different focus of the databases and partly reflects the incompleteness of the curation effort, strongly advocating collaborative initiatives in this 'curation domain' such as the ones implemented in IMEx for the curation of physical interactions. Moreover, differently from PPI, in the signalling interaction curation field no common curation policy and standards have been agreed upon. This implies that different resources often annotate the same information in different ways.

SIGNOR aims at combining curation quality, depth of annotation details, data accessibility and data usability. Moreover thanks to the annotation of the effect of single modified residues on the activity of the target protein, SIGNOR is a useful tool to map the results of high throughput functional experiments such as high throughput differential phospho-proteomics. To summarize SIGNOR stores more than 8200 enzymatic reactions where the catalytic activity of the upstream enzyme modulates the activation status or the stabilization of the downstream one. In approximately 70% (5946) of the relationships it was possible to recover the information about the modified residue whose modification cause activation or inhibition. The vast majority of the modifications are phosphorylation reactions (94%, of which 2089 have an activation effect while 3248 are inhibitory).

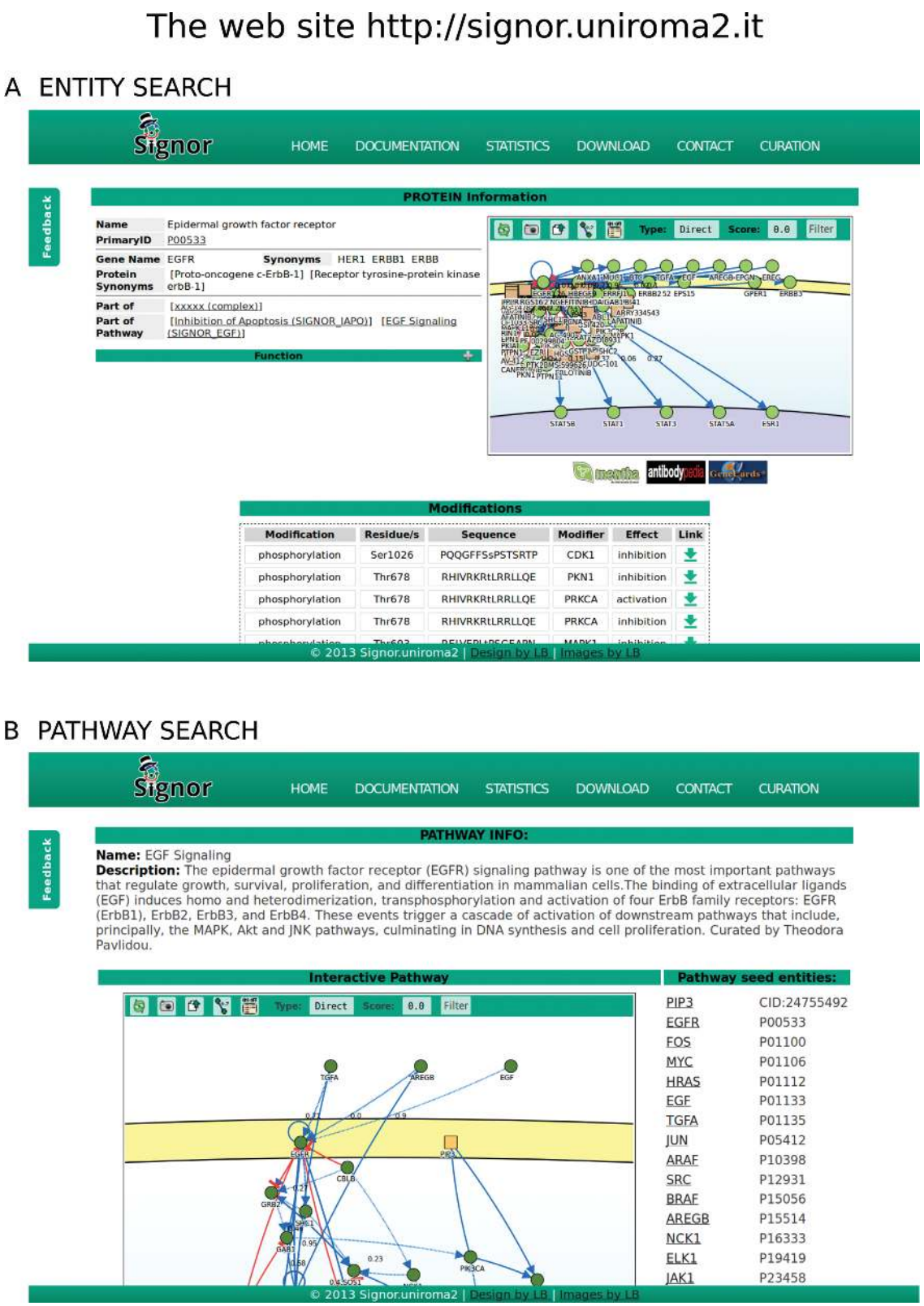

Figure 2. SIGNOR, website. The screenshot in (**A**) is an example of a result of an 'entity search' (EGFR). The web page is organized in four parts: the entity information summary, the graphic visualizer, the list of regulatory post-translational modification (reporting the modifier and the effect that the modification has on the host protein) and the list of logic relationships (not shown) that involve the query entity. The screenshot in (**B**) Represents an example of a 'pathway search' (EGF signalling), the web page is organized in three parts: the pathway description, the interactive graphic visualizer and the editable list of pathway 'seeds'. The attributes if nodes and edges are represented with different colours and symbols.

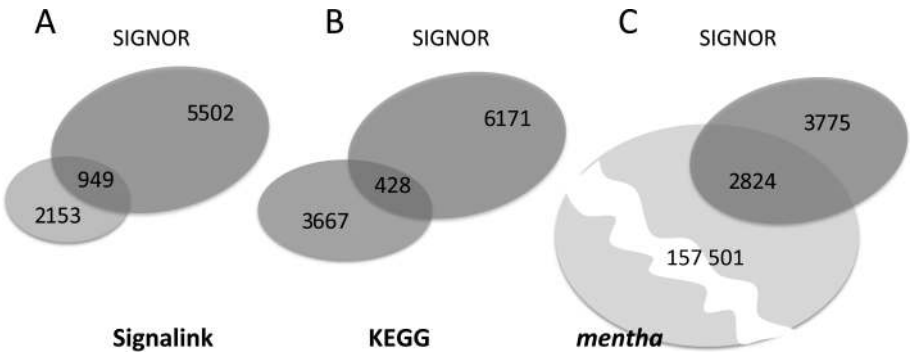

Figure 3. Comparison of SIGNOR data set with that of other databases. The Venn diagrams in (**A**), (**B**) and (**C**) represent the comparison of the coverage and the overlap in the data sets of SIGNOR and three other databases. In this comparison only the direct interactions between two proteins are considered. Interactions involving RNAs, small molecules chemicals, stimuli or phenotypes have been excluded. For SignaLink only the 'manually curated' interactions have been considered (9). KEGG archives manually curated biochemical (metabolic pathways) and causal interactions (8). In order to compare it with SIGNOR we considered all causal interactions archived in the following sub categories: Environmental Information Processing; Cellular Processes; Organismal Systems; Human Diseases. We downloaded all these pathways in kgml format and parsed them to extract activation and inhibition interactions. Furthermore, each entity involved in these interactions was remapped to UniProt primary identifies using UniProt services (12).

Less represented are dephosphorylation (4%) and ubiquitination reactions (0.3%). Finally the link of signalling proteins to chemical inhibitors (≈420 chemical compounds) and their activities to specific antibodies makes SIGNOR a unique resource to help in the design of new experiments and in the modelling of cell behaviour in response to stimuli.

Coverage and perspectives

The primary goal of SIGNOR is to provide a platform for the interpretation of experimental data and for the design of new experiments. Although incomplete, SIGNOR has a good coverage of signalling proteins and signalling interactions: PathCards is a recently developed resource integrating signalling information from multiple pathway databases (22). We observe that SIGNOR covers approx. 22.4% of the total number of human genes annotated in PathCards. Moreover, the gene products that are missing in SIGNOR and are present in PathCards participate in biological processes related to metabolic processes, to specific diseases or to the assembly of macrostructures such as ribosomes, which are outside the scope of SIGNOR. To test the coverage of SIGNOR in areas of biological and medical interest, we checked the extent to which SIGNOR covers the network of oncological interest. Of 562 onco-related genes listed in 'the cancer Gene Census', an ongoing effort to catalogue genes whose mutations are causally related to cancer (23), 446 (79%) are annotated in SIGNOR. We made an effort to link the remaining 116 to the SIGNOR signalling information but we could not recover supporting experimental information from the literature.

SIGNOR aims at increasing the database coverage with as little bias as possible. However, there are some thematic areas that will be the focus of the curation effort over the next few years. We plan to extend the coverage of dephosphorylation reactions, of ligand-receptors interactions and the curation of transcription factors, which constitute underrepresented areas in the database.

SIGNOR captures causal interactions that can be used to help assembling large logic models. In the present implementation no effort has been made to define the logic gates integrating multiple signals affecting the activation of a target entity. Although this type of information is sparse, sometimes contradictory and difficult to recover, the annotation of logic gates is a primary interest of the SIGNOR curation effort.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENT

We would like to thank Doron Lancet for useful discussions and for providing the gene co-citation table that was used to develop the reliability score.

FUNDING

European Research Council [N°322749 DEPTH to G.C.]; Italian Association for Cancer Research (AIRC) [to G.C.]. Funding for open access charge: European Research Council [N°322749-DEPTH].
Conflict of interest statement. None declared.

REFERENCES

1. Orchard,S. (2012) Molecular interaction databases. *Proteomics*, **12**, 1656–1662.
2. Orchard,S., Kerrien,S., Abbani,S., Aranda,B., Bhate,J., Bidwell,S., Bridge,A., Briganti,L., Brinkman,F.S., Cesareni,G. *et al.* (2012) Protein interaction data curation: the International Molecular Exchange (IMEx) consortium. *Nat. Methods*, **9**, 345–350.
3. Orchard,S., Ammari,M., Aranda,B., Breuza,L., Briganti,L., Broackes-Carter,F., Campbell,N.H., Chavali,G., Chen,C., del-Toro,N. *et al.* (2014) The MIntAct project–IntAct as a common curation platform for 11 molecular interaction databases. *Nucleic Acids Res.*, **42**, D358–D363.
4. Le Novere,N. (2015) Quantitative and logic modelling of molecular and gene networks. *Nat. Rev. Genet.*, **16**, 146–158.
5. MacNamara,A., Henriques,D. and Saez-Rodriguez,J. (2013) Modeling signaling networks with different formalisms: a preview. *Methods Mol. Biol.*, **1021**, 89–105.
6. Kanehisa,M., Goto,S., Sato,Y., Kawashima,M., Furumichi,M. and Tanabe,M. (2014) Data, information, knowledge and principle: back to metabolism in KEGG. *Nucleic Acids Res.*, **42**, D199–D205.

7. Croft,D., Mundo,A.F., Haw,R., Milacic,M., Weiser,J., Wu,G., Caudy,M., Garapati,P., Gillespie,M., Kamdar,M.R. *et al.* (2014) The Reactome pathway knowledgebase. *Nucleic Acids Res.*, **42**, D472–D477.
8. Kanehisa,M. and Goto,S. (2000) KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.*, **28**, 27–30.
9. Fazekas,D., Koltai,M., Turei,D., Modos,D., Palfy,M., Dul,Z., Zsakai,L., Szalay-Beko,M., Lenti,K., Farkas,I.J. *et al.* (2013) SignaLink 2 - a signaling pathway resource with multi-layered regulatory networks. *BMC Syst. Biol.*, **7**, 7.
10. Saez-Rodriguez,J., Alexopoulos,L.G., Epperlein,J., Samaga,R., Lauffenburger,D.A., Klamt,S. and Sorger,P.K. (2009) Discrete logic modelling as a means to link protein signalling networks with functional analysis of mammalian signal transduction. *Mol. Syst. Biol.*, **5**, 331.
11. Licata,L., Briganti,L., Peluso,D., Perfetto,L., Iannuccelli,M., Galeota,E., Sacco,F., Palma,A., Nardozza,A.P., Santonico,E. *et al.* (2012) MINT, the molecular interaction database: 2012 update. *Nucleic Acids Res.*, **40**, D857–D861.
12. UniProt Consortium. (2015) UniProt: a hub for protein information. *Nucleic Acids Res.*, **43**, D204–D212.
13. Wang,Y., Suzek,T., Zhang,J., Wang,J., He,S., Cheng,T., Shoemaker,B.A., Gindulyte,A. and Bryant,S.H. (2014) PubChem BioAssay: 2014 update. *Nucleic Acids Res.*, **42**, D1075–D1082.
14. Gene Ontology Consortium. (2015) Gene Ontology Consortium: going forward. *Nucleic Acids Res.*, **43**, D1049–D1056.
15. Gremse,M., Chang,A., Schomburg,I., Grote,A., Scheer,M., Ebeling,C. and Schomburg,D. (2011) The BRENDA Tissue Ontology (BTO): the first all-integrating ontology of all organisms for enzyme sources. *Nucleic Acids Res.*, **39**, D507–D513.
16. Calderone,A., Castagnoli,L. and Cesareni,G. (2013) mentha: a resource for browsing integrated protein-interaction networks. *Nat. Methods*, **10**, 690–691.
17. Dinkel,H., Chica,C., Via,A., Gould,C.M., Jensen,L.J., Gibson,T.J. and Diella,F. (2011) Phospho.ELM: a database of phosphorylation sites–update 2011. *Nucleic Acids Res.*, **39**, D261–D267.
18. Hornbeck,P.V., Zhang,B., Murray,B., Kornhauser,J.M., Latham,V. and Skrzypek,E. (2015) PhosphoSitePlus, 2014: mutations, PTMs and recalibrations. *Nucleic Acids Res.*, **43**, D512–D520.
19. Ben-Shlomo,I., Yu Hsu,S., Rauch,R., Kowalski,H.W. and Hsueh,A.J. (2003) Signaling receptome: a genomic and evolutionary perspective of plasma membrane receptors involved in signal transduction. *Sci. STKE*, **2003**, RE9 .
20. Vaquerizas,J.M., Kummerfeld,S.K., Teichmann,S.A. and Luscombe,N.M. (2009) A census of human transcription factors: function, expression and evolution. *Nat. Rev. Genet.*, **10**, 252–263.
21. Liberti,S., Sacco,F., Calderone,A., Perfetto,L., Iannuccelli,M., Panni,S., Santonico,E., Palma,A., Nardozza,A.P., Castagnoli,L. *et al.* (2013) HuPho: the human phosphatase portal. *FEBS J.*, **280**, 379–387.
22. Belinky,F., Nativ,N., Stelzer,G., Zimmerman,S., Iny Stein,T., Safran,M. and Lancet,D. (2015) PathCards: multi-source consolidation of human biological pathways. *Database: J. Biol. Databases Curation*, **2015**, bav006.
23. Futreal,P.A., Coin,L., Marshall,M., Down,T., Hubbard,T., Wooster,R., Rahman,N. and Stratton,M.R. (2004) A census of human cancer genes. *Nat. Rev. Cancer*, **4**, 177–183.