

## Uniform Curation Protocol of Metazoan Signaling Pathways to Predict Novel Signaling Components

Máté Pálffy, Illés J. Farkas, Tibor Vellai, and Tamás Korcsmáros

### Abstract

A relatively large number of signaling databases available today have strongly contributed to our understanding of signaling pathway properties. However, pathway comparisons both within and across databases are currently severely hampered by the large variety of data sources and the different levels of detail of their information content (on proteins and interactions). In this chapter, we present a protocol for a uniform curation method of signaling pathways, which intends to overcome this insufficiency. This uniformly curated database called Signalink (<http://signalink.org>) allows us to systematically transfer pathway annotations between different species, based on orthology, and thereby to predict novel signaling pathway components. Thus, this method enables the compilation of a comprehensive signaling map of a given species and identification of new potential drug targets in humans.

We strongly believe that the strict curation protocol we have established to compile a signaling pathway database can also be applied for the compilation of other (e.g., metabolic) databases. Similarly, the detailed guide to the orthology-based prediction of novel signaling components across species may also be utilized for predicting components of other biological processes.

**Key words** Literature curation, Signaling database, Signalogs, Orthology-based prediction

---

## 1 Introduction

Signal transduction pathways, functional building blocks of intracellular signaling, control various cellular processes, including cell growth, proliferation, differentiation, and stress response in divergent animal phyla [1]. In humans, defects in intracellular signaling can cause various diseases, *such as* cancer, neurodegeneration, muscle atrophy, immune deficiency, or diabetes. Therefore, a better understanding of the structure, function, and evolution of signal transduction is important for both basic research and medicine. This requires the construction of a comprehensive signaling map, which would (ideally) contain all components of distinct signaling pathways and their genetic and physical interactions. Genome programs and high-throughput (HTP) protein–protein interaction

analyses have greatly contributed to the construction of signaling maps in various model organisms, ranging from invertebrates to mammals. Accordingly, the effort to map novel signaling components and interactions has largely benefited from network alignment techniques and other widely used functional genomics methods, allowing the integration of functional data among and within species [2, 3].

Most of these methods predict new gene or protein properties (annotations) on the basis of sequence homology and similarities between known functions. Similar annotation transfer approaches have been applied to predict structural properties (e.g., domain composition), expression profiles, and physical interactions of proteins [4–6]. For predicting interactions, several techniques have been suggested, out of which one of the most widely used is the method of “interologs”: two proteins are predicted to physically interact with each other, if their orthologs in another organism also interact [7]. Interologs, however, are found to be less conserved than orthologs [8] and also less reliable than interactions generated by HTP approaches [9].

Despite a great wealth of protein interaction data obtained from HTP experiments, such as yeast two-hybrid screens, the low abundance of extracellular, membrane-bound, and nuclear signaling components (e.g., ligands, receptors, and transcription factors) make these experimental techniques only partially efficient for identifying signaling interactions [10]. Accordingly, several signaling pathway databases have been generated manually by collecting relevant data from the literature [11]. However, so far most of them lack those key features (e.g., uniform pathway curation across more than one species) that would be necessary for transferring signaling pathway membership information between species [10]. Reliable and detailed signaling pathway databases are crucial for predicting novel signaling components because they are needed (1) as sources of known pathway information from which prediction can be performed (i.e., seed data) and (2) as reference data sets against which the novelty of predictions can be tested (i.e., those predicted signaling pathway member proteins that are already known pathway members should be removed from the list of predictions, while others can be regarded as predicted components).

A comprehensive pathway resource, Signalink, developed in our lab, applies uniform curation rules to keep the levels of detail identical in all examined pathways for *Caenorhabditis elegans*, *Drosophila melanogaster*, and humans [12]. Compared to three widely used pathway databases (KEGG, Reactome and NetPath), Signalink contains the (1) highest numbers of signaling proteins and interactions; (2) highest numbers of signaling cross-talks and multi-pathway proteins; (3) and above the average number of publications used per pathway [12]. Moreover, the uniform curation protocol and data structure of the Signalink database allow

systematic transfer of pathway annotation between two species on the basis of sequence orthology.

The topology of signaling pathways is crucial for selecting possible novel drug target candidates [13]. As an example, drugs used for inhibiting a specific signaling protein in order to affect proliferation may actually activate the corresponding pathway by triggering an unknown negative feedback loop [14]. Transferring signaling pathway annotations across species may alleviate such difficulties and can provide a more comprehensive signaling network. Identification of novel signaling components may help to discover novel drug targets as (1) these signaling components can increase the applicability of model organisms for testing drugs and drug target candidates, (2) in humans, they can serve as potential novel drug targets, and (3) in the case of already used target proteins they can help to uncover possible side effects.

## 2 Materials

1. The data serving as a basis for building the SignaLink database were obtained from both review papers and primary research articles (*see* Table 1).
2. These were complemented with data derived from species-specific databases for *Drosophila* and *C. elegans* (Flybase and Wormbase, respectively) that contain information from different sources—ranging from large-scale experiments to primary research articles (*see* Table 1).
3. We collected Ensembl IDs for human proteins from the genome browser Ensembl and ORFs for worms and flies from species-specific databases (Flybase and Wormbase), while UniProt IDs were collected from UniProt for all three species (*see* Table 1).

**Table 1**  
Sources of the manually curated SignaLink database

Source	Protein	Signaling interaction	Link	Reference
170 Review papers	✓	✓		
771 Research articles	✓	✓		
Wormbase	✓	✓	<a href="http://www.wormbase.org/">http://www.wormbase.org/</a>	[37]
Flybase	✓	✓	<a href="http://flybase.org/">http://flybase.org/</a>	[32]
UniProt	✓		<a href="http://www.uniprot.org/">http://www.uniprot.org/</a>	[29]
Ensembl	✓		<a href="http://www.ensembl.org/">http://www.ensembl.org/</a>	[24]

**Table 2**  
**Search engines used for the compilation of Signalink**

Search engines	Protein	Signaling interaction	Link	Reference
iHOP	✓	✓	<a href="http://www.ihop-net.org/">http://www.ihop-net.org/</a>	(18)
Chilibot	✓	✓	<a href="http://www.chilibot.net/">http://www.chilibot.net/</a>	(15)
PubMed		✓	<a href="http://www.ncbi.nlm.nih.gov/pubmed/">http://www.ncbi.nlm.nih.gov/pubmed/</a>	
InParanoid	✓		<a href="http://inparanoid.sbc.su.se/">http://inparanoid.sbc.su.se/</a>	(17)

- We searched directly for suggested interactions between two selected proteins with iHOP and ChiliBot [15, 16] (see Table 2). iHOP uses genes and proteins as hyperlinks between sentences and abstracts, meaning that information of a single protein and its interaction is given as a sentence retrieved from source abstracts [16].
- We also used the synonym identification tool of iHOP for collecting protein synonyms.

### 3 Methods

In this section, we describe a unified curation protocol for assigning signaling proteins to signaling pathways and for compiling signaling interactions within a pathway. This standardized curation protocol in three different organisms is a prerequisite for enabling systematic transfer of pathway annotations between different species to predict new signaling components based on orthology.

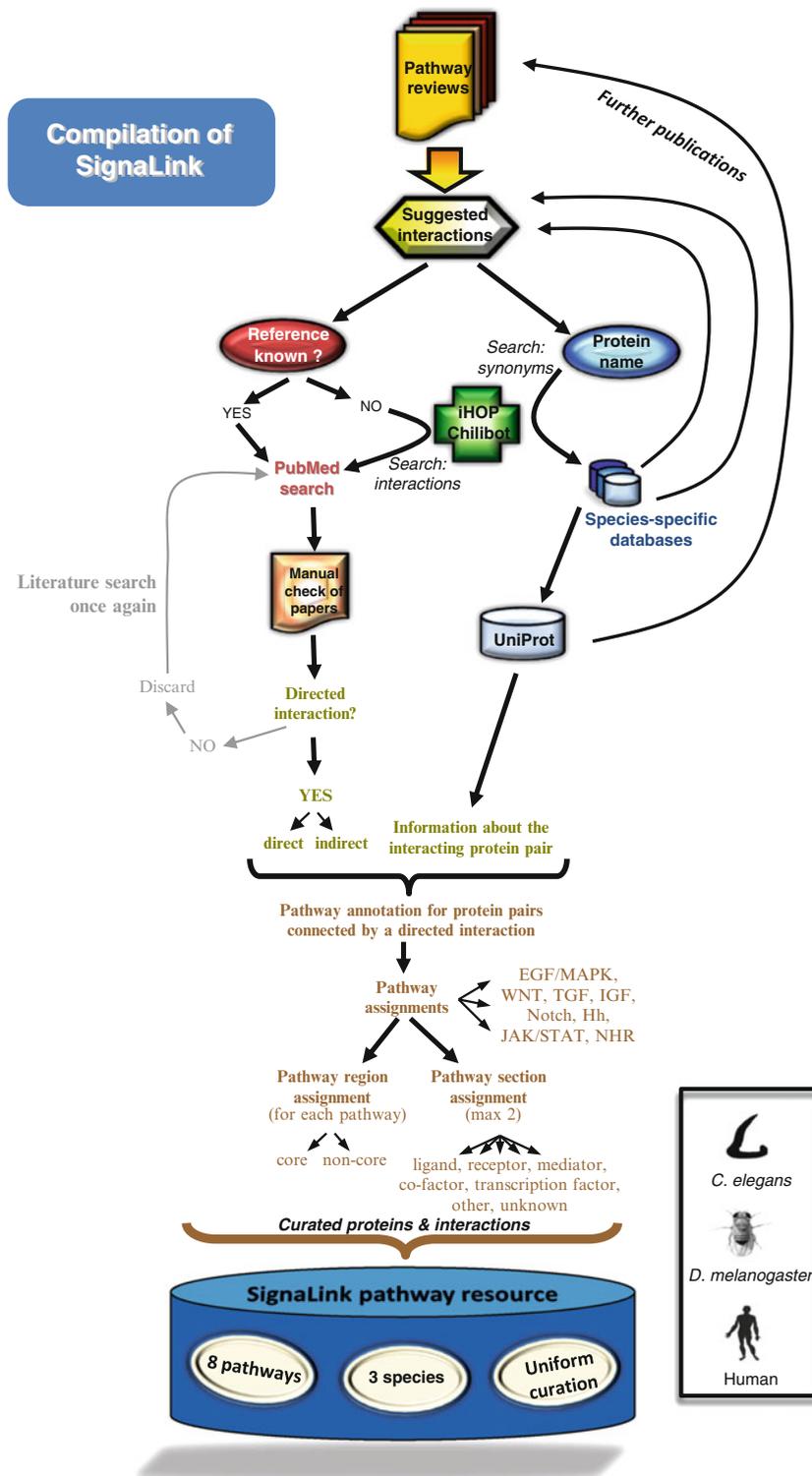
#### 3.1 Creating a Signaling Database (Signalink) by a Uniform Manual Curation Protocol

The following section describes our workflow for the construction of a signaling database, which contains eight pathways in three species (see Fig. 1). The main steps involve listing signaling proteins of the given pathways, collecting information on the proteins, assigning each protein to the region/section of a given pathway, and collecting protein interaction information of the proteins, thereby also compiling additional proteins to the pathway.

##### 3.1.1 Collecting Pathway Information for Signaling Proteins

All pathways examined from three species (*C. elegans*, *D. melanogaster*, and *Homo sapiens*) were compiled (i.e., manually curated) separately. For the challenges and importance of pathway definitions, see **Note 1**. For each pathway, three main steps were performed:

- A search for pathway-specific review articles and databases using PubMed, Google Scholar, and Google.
- The assignment of signaling proteins to signaling pathways based on the full text of reviews.



**Fig. 1** Manual curation process of Signalink. To compile the Signalink pathway resource [12] (<http://signalink.org>), signaling interactions were collected from pathway reviews, species-specific databases, and UniProt. Only interactions with references were included after manual checks via PubMed. The iHOP and ChiliBot search engines were used for finding references for suggested interactions lacking a reference in the reviews, and these search results were also manually checked. Synonyms for the interacting proteins were obtained with the help of the synonym finder tool of iHop. Finally, curated signaling proteins were assigned to pathway regions and pathway sections

3. An extended search for additional pathway proteins using iHOP and ChiliBot [15, 16].
4. When inserting a protein into SignalLink we assigned it to one pathway and—within this pathway—to one pathway region. Later, further pathways and pathway regions were added for this protein, if necessary. We marked a protein as a “core” component of a pathway, if it is essential for transmitting the signal of its pathway and has at least one of the pathway’s biochemical characteristics, e.g., “Ser/Tyr-kinase activity”. A “non-core” (or “peripheral”) component modulates the pathway’s core proteins, but it does not participate directly in the transduction of the signaling flow.
5. Additionally, the pathway section(s) of each protein was determined separately and a maximum of two sections per protein were allowed. The pathway position *ligand* indicates that the given protein initiates the signal of its pathway. A *receptor* is the direct receiver of this signal. A *mediator* is a member of the pathway that transduces the signal from the receptor towards downstream transcription factors. A *co-factor* modulates the function of any other protein from the pathway. Notably, co-factors often reside in the peripheral (non-core) region of their pathways. A *transcription factor* (1) activates another transcription factor (TF) after receiving the signal from its pathway, or (2) forms a complex with other TF proteins, or (3) binds to a specific promoter region (i.e., a specific binding site) on the DNA. Non-signaling proteins with roles in cellular motion, transport, and membrane anchoring were marked as *other*. When information on the position of a signaling protein in its pathway was lacking, the protein was marked *unknown*.

### 3.1.2 Collecting Signaling Protein Information

After listing pathway proteins from review and research papers, information on the signaling proteins were collected from different databases (*see* Table 1). For each protein, we also listed its orthologs in the other two species with the help of the ortholog clusters of the InParanoid database [17]. During collecting UniProt IDs, if more than one UniProt ID were available for the same protein, then the ID(s) of the protein(s) with the longest amino acid sequence was (were) used. To make the database more comprehensive, we assigned all known synonyms of the proteins. These were listed from review papers, and the “synonym” field of the iHOP database [18]. For the conversion of protein IDs, the Protein Identifier Cross-Reference Service (PICR) [19] and Synergizer [20] were used.

### 3.1.3 Collecting Signaling Interaction Information

A key feature of a signal transduction network is that the direction of an interaction is well distinguishable (e.g., protein A activates or negatively regulates protein B). Accordingly, all interactions

inserted into SignaLink had to be directed. Each interaction had to be documented with the PubMed ID of the publication reporting the verifying experiment(s). Signaling interactions of a protein were collected from primary research articles, listed in review papers, species-specific databases (FlyBase, WormBase), and UniProt, iHOP, ChiliBot, and PubMed search results (*see* Tables 1 and 2). All research articles were manually examined, and in terms of biochemical experimental evidence, we marked every protein interaction as either *direct* or *indirect*. Direct experimental evidence indicates that there is a published biochemical evidence for signaling interaction between two given proteins, whereas indirect experimental evidence indicates that there is no direct biochemical evidence for interaction, but published experimental results suggest that interaction is very likely possible. Evidence types accepted here involve (1) changes in mRNA/protein levels, enzyme activities, concentrations of the products of catalyzed reactions, and (2) docking domain structures.

Importantly, not only the direction, but also the effect of an interaction is highly relevant to a signaling database. All interactions can be characterized as *activating* or *inhibitory*.

For interactions with indirect evidence, we marked activating interactions as ++ and --, while inhibitory interactions were marked +- and -+. A unidirectional interaction (A and B interact as either A→B or B→A) has only one type of effect, but for the few bidirectional interactions (A→B and B→A are both present) more than one type of effects are possible between the two proteins. Two signaling interactions between the same two proteins in opposing directions are listed separately in SignaLink. For the challenges and limitations of manual curation, *see* **Note 2**.

### 3.1.4 Curation Process

*Example: The Notch Signaling Pathway and the NOTCH1 Protein*

As an example, we present here the human Notch pathway and one of its components, the human NOTCH1 receptor protein. We describe the process of (1) obtaining information for the protein NOTCH1 and (2) obtaining protein interaction information for NOTCH1.

According to pathway-based reviews [21], there are 4 members of Notch receptor family proteins in humans: NOTCH1, NOTCH2, NOTCH3, and NOTCH4. Notch proteins have a specific role in transmitting signals [22] between ligands and transcription factors, as well as several additional proteins which influence the function of Notch proteins [23].

Alternative splicing can generate functionally different proteins from the same coding region; however, in the majority of proteins functional significance of different splice variants remains unknown. Despite their potentially different roles, databases and review papers do not differentiate between splice variants. For the human NOTCH1 protein, Ensembl [24] contains two splice variants: ENSP00000277541 and ENSP00000360765. From these

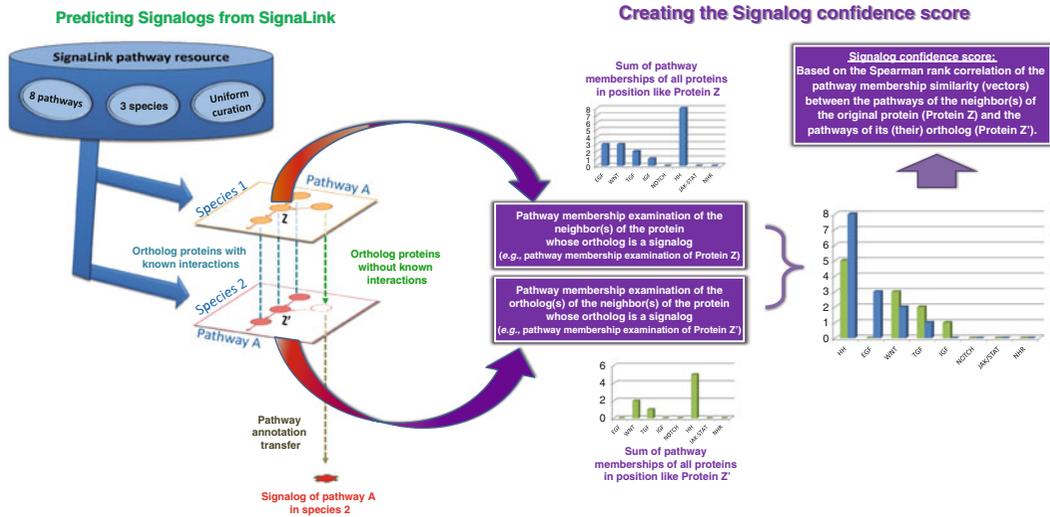
two, the InParanoid database [17] contains only the first, ENSP00000277541. Therefore, we inserted only this splice variant into SignaLink. For proteins that have more than one splice variant, but none of them is present in the InParanoid database, we inserted into SignaLink the splice variant that has a primary UniProt accession (AC), as listed by Ensembl version 49.

From Ensembl, we included into SignaLink the UniProt accession(s) of a protein, and from UniProt we used the following data fields of the protein: description, reference—if it contained interaction data,—and cellular component. In addition, data from protein description and interaction fields were manually tested in primary publications for further information.

Regarding the region/section, NOTCH1 is a core protein of its pathway and functions as a receptor, mediator, or transcription factor, according to ref. [25]. However, within the Notch pathway, NOTCH1 functions either as a receptor or a transcription factor. Thus, we included only these two pathway sections for NOTCH1 into SignaLink.

To make SignaLink as complete as possible, we searched for orthologs of the human NOTCH1 protein. Orthologs without known signaling interactions became predicted pathway proteins in SignaLink. From the InParanoid database we identified the *C. elegans* and *D. melanogaster* orthologs of human NOTCH1 (ENSP00000277541). (In several cases we searched by both the UniProt and Ensembl protein IDs in InParanoid to find the protein.) Interestingly, the human NOTCH1 has two worm orthologs (LIN-12 and GLP-1), but only one fly ortholog (the protein N). We inserted all three orthologs into SignaLink. We listed species-specific protein IDs and UniProt ACs of the orthologs from WormBase and FlyBase. For ligands and transcription factors interacting with NOTCH1, we followed the same steps.

Next, we listed articles describing signaling interactions between NOTCH1 and other proteins by browsing through the references of the above mentioned review papers and by using the search engines iHOP [18] and ChiliBot [15]. iHOP allows users to search for all abstracts with interactions containing NOTCH1. With ChiliBot the interaction between two selected proteins can be directly searched for. As an example, interaction between NOTCH1 and TACE/ADAM17 has been described in an experimental article [26]. After reading the article, we found that it describes (1) a putative cleavage site for TACE on NOTCH1 and (2) a correlation between the in vitro enzymatic activity of TACE and the activity of NOTCH1. Thus, this article provides evidence for the activation of NOTCH1 by TACE. In addition, we directly searched for interactions between the orthologs of TACE and NOTCH1 in the other two species.



**Fig. 2** Prediction of signalogs and calculation of the signalog confidence score. Based on the Signalink resource [12] orthology assignment was performed between each pair of the three species. Proteins were predicted to be members of the same signaling pathway(s) where their orthologs belong. An interaction with a signaling protein  $Z'$  was predicted for a protein, if the ortholog of the protein interacted with  $Z$  (the ortholog of  $Z'$ ) in the same pathway  $A$  in a different species. A confidence score was calculated based on the pathway membership similarity between the neighbors of  $Z$  and its ortholog  $Z'$ . See main text for details

### 3.2 Signalog Prediction Based on Orthologous Signaling Components

Despite the conservation of many biological processes (e.g., developmental signaling pathways) throughout evolution, there is a poor overlap in protein–protein interactions between species in different databases [27]. Furthermore, the catalogue of proteins annotated with signaling function is incomplete even in highly studied model organisms. Therefore, the prediction of new potential signaling interactions and also new signaling proteins based on orthology is an important task.

#### 3.2.1 Prediction of Signalogs

We started with creating a list from the three species examined in Signalink (*C. elegans*, *D. melanogaster*, and *H. sapiens*) by collecting those proteins that have no known signaling interactions, but have at least one signaling pathway member ortholog in the other two species. Similarly to the concept of functional orthology [28], for each of these proteins we assumed that their pathway annotations (i.e., signaling role) can be transferred between species. Thus, we predicted that a protein is a member of the same signaling pathway(s) in which its ortholog(s) belong(s) (see Fig. 2). These proteins were termed as signalog proteins (signalogs). Because in Signalink a protein can belong to more than one pathway [12], a signalog can also be annotated to more than one pathway. Using this approach we were able to predict 88, 92, and 73 novel signaling proteins in worms, flies, and humans, respectively [10]. For the limitations of orthology-based pathway annotation transfer, see **Note 3**.

### 3.2.2 Defining the Novelty of Signaling Protein Predictions Based on Orthology

To verify the novelty of the predicted signaling roles which have not been featured in other resources yet, we searched the literature with semiautomated methods for already known annotations. Next, we compared the list of signalogs and their predicted pathway memberships to pathway annotations found in pathway databases, as well as the list of ortholog predictions to previously published interolog predictions. To assess the novelty of signalogs and quantify the confidence level of each prediction, we performed semiautomated searches using PubMed, UniProt, GO, Wormbase, FlyBase, iHOP, and Chilibot web services [15, 18, 29–32]. During this process, direct manual curation and Python scripts checking multiple proteins in one webservice were used. In each of the three species examined, we classified the predicted signalogs into five groups on the basis of their known properties in the literature: (1) no orthology information and/or no biochemical function is available; (2) there are known orthologs with unknown biochemical function; (3) only biochemical function is available, but orthology information is lacking; (4) data on orthology as well as biochemical function(s) exist; (5) orthologs, biochemical function(s), and pathway annotation(s) are all known. Categories 1–5 denote a decreasing level of novelty. However, even category (5) contains signalogs for which at least one novel signaling pathway membership is predicted. Additionally, to check the novelty of the predicted signaling pathway memberships, we compared the list of signalogs and their predicted pathway memberships to known pathway membership annotations from Reactome and KEGG [33, 34]. We next applied interologs to verify the novelty of our ortholog predictions (an interolog is a pair of proteins predicted to interact based on the interaction of the two proteins' orthologs in at least one other organism) [7]. To reveal the presence of signalogs in current orthology-based prediction databases, we compared already identified interologs in worms, flies, and humans using three species-specific datasets (WI8, DroID, and HomoMINT) [8, 35, 36] with interologs generated from SignaLink data. Since neither SignaLink [12] nor the current signalog identification approach identify interologs directly, we used an indirect method by first deducing interologs from SignaLink data: we linked two proteins in an organism, if their orthologs interacted in at least one of the other three organisms. After generating all possible interologs from SignaLink, we examined only those (predicted interactions) in which at least one of the interactors is a signalog protein (predicted signaling pathway member).

### 3.2.3 Creating a Confidence Score for Signalogs

To assess the reliability of a signalog, a confidence score was calculated in each case (*see* Fig. 2). For the signalog  $Z'$  that was predicted to be a component of Pathway  $A'$  ( $PA'$ ) in Species 2, we examined pathway membership of each neighbors (protein interactors) of  $Z'$

in Species 2 and the known signaling component,  $Z$  in Species 1. For each  $Z$  and  $Z'$  proteins, we summed pathway memberships as 2 pathway vectors (Vector\_ $Z$  and Vector\_ $Z'$ ). Vectors have components indexed with the name of their signaling pathways. Finally, we computed the Spearman rank correlation of vectors computed for  $Z$  and  $Z'$ , and based on this correlation we defined the Signalog confidence score:  $[(\text{Spearman\_corr} + 1)/2] * 100$ . This confidence score quantifies similarity between the signaling pathway membership profile of the possible interactors of a signalog protein and the original signaling protein (i.e., the orthologs of the signalog protein). Predictions above 50% can be considered as confident predictions.

---

## 4 Notes

1. Pathway definition is a critical task when compiling a pathway database. Pathway databases tend to use different pathway definitions, such as:
  - Canonical (e.g., MAPK)
  - Functional (e.g., inflammation)
  - Inferred (e.g., from gene expression data)
  - Cellular process regulating (e.g., autophagy induction)
  - Organ-related (e.g., vulva development)
  - Disease-related (e.g., list of connected proteins affected by mutations in breast cancer; Alzheimer's disease)
  - Drug-related (e.g., pharmacologically affected list of connected proteins)

To develop a database for comparative purposes or systems-level examinations, pathway definitions must be the same in the whole database. For Signalog, we applied a biochemically based, well-documented, and clear pathway definition. For example, the EGF/MAPK pathway in Signalog contains (with evolutionary and biochemical reasoning) the pathway from the EGF ligand to the terminal MAPK kinases. In several other databases this pathway is scattered across many separate (sub)pathways (e.g., EGFR, RAS, p38, JNK, ERK, ASK). An important consequence of precise pathway definitions is the reduced number of examined pathways. An appropriate and precise grouping can be important to avoid artificial pathway constructs [12].

2. Despite recent advances in the technology of manual pathway curation, this technology still does have several limitations. First, curation highly depends on the knowledge and background of the curator as well as on the quality of the protocol

used for the curation [9]. Second, all data are based on the actual knowledge from the literature. Therefore, these databases have to be updated regularly (e.g., annually or bi-annually).

3. According to our current knowledge, the limitations of systematical pathway annotation transfer between species are the following. Interactions of membrane-bound and nuclear proteins are still underrepresented in most databases, thus predictions involving these proteins are less reliable. Furthermore, interactions between signaling proteins have been shown to be generally more unique to their species than PPIs in most biological processes [7].

---

## Acknowledgement

Authors were supported by the European Union and the European Social Fund [TAMOP-4.2.1/B-09/1/KMR-2010-0003], the Hungarian Scientific Research Fund [OTKA K75334, NK78012], and a János Bolyai Scholarship to TK and TV.

## References

1. Pires-daSilva A, Sommer RJ (2003) The evolution of signalling pathways in animal development. *Nat Rev Genet* 4:39–49
2. Gabaldon T, Huynen MA (2004) Prediction of protein function and pathways in the genome era. *Cell Mol Life Sci* 61:930–944
3. Kuzniar A, van Ham RC, Pongor S et al (2008) The quest for orthologs: finding the corresponding gene across genomes. *Trends Genet* 24:539–551
4. Yellaboina S, Dudekula DB, Ko MS (2008) Prediction of evolutionarily conserved interologs in *Mus musculus*. *BMC Genomics* 9:465
5. Storm CE, Sonnhammer EL (2003) Comprehensive analysis of orthologous protein domains using the HOPS database. *Genome Res* 13:2353–2362
6. Salgado D, Gimenez G, Coulier F et al (2008) COMPARE, a multi-organism system for cross-species data comparison and transfer of information. *Bioinformatics* 24:447–449
7. Yu H, Luscombe NM, Lu HX et al (2004) Annotation transfer between genomes: protein-protein interologs and protein-DNA regulogs. *Genome Res* 14:1107–1118
8. Persico M, Ceol A, Gavrila C et al (2005) HomoMINT: an inferred human network based on orthology mapping of protein interactions discovered in model organisms. *BMC Bioinformatics* 6(Suppl 4):S21
9. Cusick ME, Yu H, Smolyar A et al (2009) Literature-curated protein interaction datasets. *Nat Methods* 6:39–46
10. Korcsmaros T, Szalay MS, Rovó P et al (2011) Signalogs: orthology-based identification of novel signaling pathway components in three metazoans. *PLoS One* 6:e19240
11. Bauer-Mehren A, Furlong LI, Sanz F (2009) Pathway databases and tools for their exploitation: benefits, current limitations and challenges. *Mol Syst Biol* 5:290
12. Korcsmaros T, Farkas IJ, Szalay MS et al (2010) Uniformly curated signaling pathways reveal tissue-specific cross-talks and support drug target discovery. *Bioinformatics* 26:2042–2050
13. Chaudhuri A, Chant J (2005) Protein-interaction mapping in search of effective drug targets. *Bioessays* 27:958–969
14. Sergina NV, Rausch M, Wang D et al (2007) Escape from HER-family tyrosine kinase inhibitor therapy by the kinase-inactive HER3. *Nature* 445:437–441
15. Chen H, Sharp BM (2004) Content-rich biological network constructed by mining PubMed abstracts. *BMC Bioinformatics* 5:147

16. Hoffmann R, Valencia A (2004) A gene network for navigating the literature. *Nat Genet* 36:664
17. Berglund AC, Sjolund E, Ostlund G et al (2008) InParanoid 6: eukaryotic ortholog clusters with inparalogs. *Nucleic Acids Res* 36:D263–D266
18. Fernandez JM, Hoffmann R, Valencia A (2007) iHOP web services. *Nucleic Acids Res* 35:W21–W26
19. Cote RG, Jones P, Martens L et al (2007) The Protein Identifier Cross-Referencing (PICR) service: reconciling protein identifiers across multiple source databases. *BMC Bioinformatics* 8:401
20. Berriz GF, Roth FP (2008) The Synergizer service for translating gene, protein and other biological identifiers. *Bioinformatics* 24:2272–2273
21. Baron M (2003) An overview of the Notch signalling pathway. *Semin Cell Dev Biol* 14:113–119
22. Weinmaster G (1997) The ins and outs of notch signaling. *Mol Cell Neurosci* 9:91–102
23. Bray SJ (2006) Notch signalling: a simple pathway becomes complex. *Nat Rev Mol Cell Biol* 7:678–689
24. Flicek P, Aken BL, Beal K et al (2008) Ensembl 2008. *Nucleic Acids Res* 36:D707–D714
25. Ilagan MX, Kopan R (2007) SnapShot: notch signaling pathway. *Cell* 128:1246
26. Brou C, Logeat F, Gupta N et al (2000) A novel proteolytic cleavage involved in Notch signaling: the role of the disintegrin-metalloprotease TACE. *Mol Cell* 5:207–216
27. Gandhi TK, Zhong J, Mathivanan S et al (2006) Analysis of the human protein interactome and comparison with yeast, worm and fly interaction datasets. *Nat Genet* 38:285–293
28. Bandyopadhyay S, Sharan R, Ideker T (2006) Systematic identification of functional orthologs based on protein network comparison. *Genome Res* 16:428–435
29. Boutet E, Lieberherr D, Tognolli M et al (2007) UniProtKB/Swiss-Prot: the manually annotated section of the UniProt Knowledge-Base. *Methods Mol Biol* 406:89–112
30. Ashburner M, Ball CA, Blake JA et al (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* 25:25–29
31. Harris TW, Antoshechkin I, Bieri T et al (2010) WormBase: a comprehensive resource for nematode research. *Nucleic Acids Res* 38:D463–D467
32. Drysdale R (2008) FlyBase: a database for the *Drosophila* research community. *Methods Mol Biol* 420:45–59
33. Ogata H, Goto S, Sato K et al (1999) KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* 27:29–34
34. Joshi-Tope G, Gillespie M, Vastrik I et al (2005) Reactome: a knowledgebase of biological pathways. *Nucleic Acids Res* 33:D428–D432
35. Yu J, Pacifico S, Liu G et al (2008) DroID: the *Drosophila* Interactions Database, a comprehensive resource for annotated gene and protein interactions. *BMC Genomics* 9:461
36. Simonis N, Rual JF, Carvunis AR et al (2009) Empirically controlled mapping of the *Caenorhabditis elegans* protein-protein interactome network. *Nat Methods* 6:47–54
37. Rogers A, Antoshechkin I, Bieri T et al (2008) WormBase 2007. *Nucleic Acids Res* 36:D612–D617