Uniformly curated signaling pathways reveal tissue-specific cross-talks and support drug target discovery

Tamás Korcsmáros1,2,†, Illes J. Farkas3,†, Máté S. Szalay2, Petra Rovó1, Dávid Fazekas1, Zoltán Sipőcz2, Csaba Böde4, Katalin Lenti5, Tibor Vellai1 and Péter Csermely2,∗

1Department of Genetics, Eötvös University, Pázmány P. s. 1C, H-1117 Budapest, 2Department of Medical Chemistry, Semmelweis University, PO Box 260, H-1444 Budapest, 3Statistical and Biological Physics Group of the Hungarian Academy of Sciences, Pázmány P. s. 1A, H-1117 Budapest, 4Morgan Stanley Hungary Analytics Ltd., Loschner Ö. f. 8, H-1085 Budapest and 5Department of Morphology and Physiology, Semmelweis University, Vas u. 17, H-1088 Budapest, Hungary

Associate Editor: Jonathan Wren

ABSTRACT

Motivation: Signaling pathways control a large variety of cellular processes. However, currently, even within the same database signaling pathways are often curated at different levels of detail. This makes comparative and cross-talk analyses difficult.

Results: We present SignalLink, a database containing eight major signaling pathways from Caenorhabditis elegans, Drosophila melanogaster and humans. Based on 170 review and ~800 research articles, we have compiled pathways with semi-automatic searches and uniform, well-documented curation rules. We found that in humans any two of the eight pathways can cross-talk. We quantified the possible tissue- and cancer-specific activity of cross-talks and found pathway-specific expression profiles. In addition, we identified 327 proteins relevant for drug target discovery.

Conclusions: We provide a novel resource for comparative and cross-talk analyses of signaling pathways. The identified multi-pathway and tissue-specific cross-talks contribute to the understanding of the signaling complexity in health and disease, and underscore its importance in network-based drug target selection.

Availability: http://signalink.org

Contact: csermely@eok.sote.hu

Supplementary information: Supplementary data are available at Bioinformatics online.

Received on March 1, 2010; revised on April 29, 2010; accepted on June 4, 2010

1 INTRODUCTION

Intracellular signaling, from the simplest cascades to the highly intertwined networks of protein kinases, contributes extensively to the diversity of developmental programs and adaptation responses in metazoans (Pires-da-Silva and Sommer, 2003). In humans, defects in intracellular signaling can cause various diseases, e.g. cancer, neurodegeneration or diabetes. Thus, understanding the structure, function and evolution of signal transduction is an important task for both basic research and medicine. By now genetic studies have uncovered functionally separate, though interacting (cross-talking), pathways and the direction of information flow between pairs of signaling molecules in a number of species (Beyer et al., 2007). On the other hand, biochemical experiments have allowed the detailed characterization of direct physical interactions involved in signaling (Xia et al., 2004). Integrating these data sets using uniform manual curation criteria can significantly contribute to a more precise assessment of their tissue- and cancer-specific utilization and the effects of drug treatments (Davidov et al., 2003). For example, inhibitors used for eliminating a signaling pathway in cancerous cells may in fact have the opposite effect. These drugs may suppress negative feedback loops and thereby, paradoxically, activate the targeted pathway (Sergina et al., 2007).

Intracellular signaling was originally regarded as an assembly of distinct and almost linear cascades. Over the past decade, however, it has been realized that signaling pathways are highly structured and rich in cross-talks (where cross-talk is defined here as a directed physical interaction between pathways). Consequently, intracellular signaling is now viewed as a set of intertwined pathways forming a single signaling network (Papin et al., 2005). This paradigm shift calls for novel experimental, curation and network modeling techniques (Bauer-Mehren et al., 2009).

Currently, high-throughput (HTP) experiments are the major sources of known protein-protein interactions (PPIs). However, so far in most HTP experiments extracellular, membrane-bound and nuclear proteins have been underrepresented. These and other sampling biases strongly reduce their usability for identifying signaling interactions. Another limitation of HTP assays is that they produce undirected interactions even though in signaling directions are essential. Accordingly, several signaling pathway databases have been created recently by manually collecting the directed interactions from the literature (Bauer-Mehren et al., 2009).

Manually curated signaling pathway databases are often assembled without strictly defined and published standardized curation criteria (Lu et al., 2007). Therefore, even within the same database, e.g. in KEGG (Kyoto Encyclopedia of Genes and Genomes; Ogata et al., 1999), the level of detail of curation and the rules for setting pathway boundaries can vary among pathways. In addition, in several signaling resources the definition of signaling pathways has no evolutionary or biochemical background. In other cases, e.g. in Reactome and NetPath (Joshu-Tope et al., 2005;
Another limitation of several current signaling resources is that they neglect the importance of multi-pathway proteins, i.e. proteins functioning in more than one pathway (Komarova et al., 2005). In summary, the manual curation process needs to be uniform across all pathways and species to aid cross-talk analyses, tests of evolutionary hypotheses, dynamical modeling, setting up predictions and drug target selection (Table 1).

We present SignaLink, a signaling resource compiled by applying uniform manual curation rules and data structures across eight major, biochemically defined signaling pathways in three metazoans (Fig. 1). The curation method allowed a systematic comparison of pathway sizes and cross-talks. We found that in humans any two (Fig. 1). The curation protocol of SignaLink (Fig. 1A) contains several steps aimed to aid cross-talk analyses, tests of evolutionary hypotheses, dynamical modeling, setting up predictions and drug target selection (Table 1).

We present SignaLink, a signaling resource compiled by applying uniform manual curation rules and data structures across eight major, biochemically defined signaling pathways in three metazoans (Fig. 1). The curation method allowed a systematic comparison of pathway sizes and cross-talks. We found that in humans any two pathways can cross-talk, and in humans we compared the possible dynamic activities of both the pathways and their cross-talks. We characterized tissue- and cancer-specific expression profiles, and identified proteins relevant for drug target discovery.

2 SYSTEM AND METHODS
2.1 Signaling proteins and interactions
SignaLink lists signaling proteins and directed signaling interactions between pairs of proteins in healthy cells of Caenorhabditis elegans, Drosophila melanogaster and Homo sapiens. Each interaction is documented with the PubMed ID of the publication reporting the verifying experiment(s). SignaLink was compiled separately for all pathways of the three organisms. Search functions, data and network images of the pathways are available at http://SignaLink.org.

In each of the three organisms, we first listed signaling proteins and interactions from reviews (and from WormBook in C.elegans) and then added further signaling interactions of the listed proteins. To identify additional interactions in C.elegans, we examined all interactions (except for transcription regulation) of the signaling proteins listed in WormBase (Rogers et al., 2008) and added only those to SignaLink that we could manually identify in the literature as an experimentally verified signaling interaction. For D.melanogaster, we added to SignaLink those genetic interactions from FlyBase (Drysdale, 2008) that were also reported in at least one yeast-2-hybrid experiment. For humans, we manually checked the reliability and directions for the PPIs found with the search engines iHipp and Chilbot (Chen and Sharp, 2004; Hoffmann and Valencia, 2004).

SignaLink assigns proteins to signaling pathways using the full texts of pathway reviews (written by pathway experts). While most signaling resources consider 5-15 reviews per pathway, SignaLink uses a total of 170 review papers, i.e. more than 20 per pathway on average. Interactions were curated from a total of 941 articles (PubMed IDs are available at the website). We added a small number of proteins based on InParanoid ortholog clusters (Berghlund et al., 2008). For curation, we used a self-developed graphical tool and PerlPython scripts. The current version of SignaLink was completed in May 2008 based on WormBase (version 191), FlyBase (2008.6), Ensembl (49), UniProt (87) and the publications listed on the website. Pathway data can be downloaded in several formats: SQL, CSV, XLS, CYS and SVG exported from CytoScape and SBML.

2.2 Quality control, database validation and statistical significance tests
The curation protocol of SignaLink (Fig. 1A) contains several steps aimed specifically at reducing data and curation errors. We used reviews as a starting point, manually looked up interactions three times, and manually searched for interactions of known signaling proteins with no signaling interactions so far in the database. The section ‘Advantages and limitations of SignaLink’ explains validation steps and results in detail. We performed functional significance tests for each of the signaling pathways and their overlaps, i.e. multi-pathway proteins, with the ‘GO Termfinder’ toolbox (Boyle et al., 2004). We found a significant functional similarity between the functions of multi-pathway proteins and the functions of their pathways compared to the control case (functional similarity between the functions of all proteins and all pathways). Moreover, we statistically evaluated the human interactions listed in SignaLink with the PRINCESS web service (Li et al., 2008) and found that the ratio of high confidence interactions is 90.6%. Details for all statistical significance tests are available in the Supplementary Material.

2.3 Expression in selected healthy tissues and liver carcinomas
To investigate the dynamic activity of pathway interactions, we selected five healthy tissue types—colorectal, muscle, skin, liver and cardiovascular

<table>
<thead>
<tr>
<th>Table 1. Comparison of the curation processes of three manually curated databases and SignaLink</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>KEGG</strong></td>
</tr>
<tr>
<td>Source of signaling proteins</td>
</tr>
<tr>
<td>Source of interactions</td>
</tr>
<tr>
<td>Number of signaling pathways</td>
</tr>
<tr>
<td>Definitions of pathways</td>
</tr>
<tr>
<td>Pathways in one platform for cross-talk analysis</td>
</tr>
</tbody>
</table>

Only the differences from SignaLink are listed. The features setting SignaLink most clearly apart have a gray background.
tissues—and two liver carcinomas (Fig. 3, see the details for the selection process of the tissues and controls in the Supplementary Material). Protein expression data in healthy tissue types were downloaded from the eGenetics database (integrated into Ensembl). Protein expression data in two screens of liver carcinomas were obtained from Oncomine 3.6 (Rhodes et al., 2007). We considered a protein differentially expressed if the \( P \)-value of its expression in at least one of the two screens, as compared to healthy liver tissues and computed by a \( t \)-test of Oncomine, was below 0.05.

2.4 Functional annotation of drug target candidates

We collected information on the proteins that can be relevant in drug target discovery with DAVID (Database for Annotation, Visualization, and Integrated Discovery; Dennis, Jr. et al., 2003). We downloaded disease-related annotations from OMIM (Online Mendelian Inheritance in Man), genetic association database (GAD) and Orthodisease (Amberger et al., 2009; Becker et al., 2004; O’Brien et al., 2004), domain information from InterPRO (Hunter et al., 2009), and molecular function and cellular component data from Gene Ontology (GO; Harris et al., 2004).

3 RESULTS

3.1 Uniform compilation of signaling pathways in three metazoan species

We curated the signaling pathways of the nematode \textit{C. elegans}, the fruit fly \textit{D. melanogaster}, and \textit{H. sapiens}. From the wide variety of classification schemes for selecting signaling pathways (Bader et al., 2006), we followed the biochemical approach of Pires-daSilva and Sommer (2003). We selected eight major pathways for curation—EGF/MAPK, IGF/IGF-\( \beta \), WNT, Hh (Hedgehog), JAK/STAT, Notch and NHR (Nuclear Hormone Receptors)—that have central roles both in development and in normal cellular signaling.

SignaLink is a manually compiled resource integrating experimentally confirmed genetic and physical interactions from healthy tissue types. Proteins and interactions are listed without tissue-specificity and can be visualized as networks of potential interactions. Tissue- and disease-specific information can be added easily as shown in the examples below. Five combined characteristics create the unique utility of SignaLink.

(1) Pathways are biochemically defined and encompass all major developmental signaling mechanisms.

(2) A protein can belong to more than one pathway (if it does, then it is called a multi-pathway protein).

(3) Proteins are tagged with (i) the pathway(s), (ii) pathway region(s) (core and peripheral) and (iii) the pathway sections (one or two of: ligand, receptor, mediator, co-factor, transcription factor and other) they belong to.

(4) The level of detail is the same for the entire database.
(5) Interactions are directed and manually labeled with PubMed IDs (experimental evidence).

Currently, SignaLink lists 860 proteins and 237 interactions from C.elegans, 344 proteins and 233 interactions from D.melanogaster and 646 proteins with 991 interactions from humans. Similarities and differences between species and pathways are shown in Figures 1B and 1C. The database, its help pages, a detailed description of the curation process, and network visualizations of all pathways are available at http://SignaLink.org.

3.2 A large-scale view of species-specific pathway and pathway section sizes

In all three organisms, a few of the eight pathways are central and abundant. Of all proteins, 26–38% participate in the EGF/MAPK and WNT pathways, respectively. Other pathways with high protein numbers are NHR in the worm, HB and Notch in the fly, and TGF and JAK/STAT in humans. Altogether in each species 68–85% of all signaling proteins participate in these pathways and 56–70% of all cross-talks involve the EGF/MAPK, TGF or WNT pathways. C.elegans has almost identical numbers of core and peripheral proteins in each pathway (except for Notch and NHR), while in the other two species the ratio of core to peripheral proteins is around 1.5.

Pathway size differences between the three species are often related to the different environments to which the cells of these organisms have adapted. For example, ligands from the environment can easily reach the nuclei of the worm’s cells, thus, the worm’s NHR pathway is exceptionally large (58% of all signaling proteins). On the other hand, due to the large variety of signals that human cells are exposed to the human JAK/STAT pathway is oversized compared to the other two species (21% of all signaling proteins in humans versus 0% and 4% in C.elegans and Drosophila, respectively).

In all three species, EGF/MAPK and IGF have high number of mediators. However, environmental differences may affect pathway section sizes too. In C.elegans transcription factors—dominated by the NHR pathway—are the largest pathway section (59%). In the other two species, co-factors by far outnumber other pathway sections (32–42%) and in humans JAK/STAT ligands and receptors are abundant.

3.3 Multi-pathway proteins: proteins functioning in more than one signaling pathway

In C.elegans, D.melanogaster and humans, we found 6, 12 and 62 multi-pathway proteins, respectively. Within one human signaling pathway the ratio of proteins functioning in at least one other pathway varies from 5% (Notch) to 46% (IGF). Interestingly, a single protein can be even a central (i.e. core) component in more than one pathway. For example, the scaffold protein AXIN and the kinase GSK3 are both core components of more than one signaling pathway (Frame and Cohen, 2001; Luo and Lin, 2004).

We found that EGF/MAPK—the largest pathway—is the only one sharing proteins with all other pathways. On the other end of the spectrum are the Notch, JAK/STAT and NHR pathways: their proteins are contained by three or four other pathways. These differences correlate well with the numbers of pathway functions. Note also that the set of 62 human multi-pathway proteins is enriched with disease-related proteins: 45% (28) of them are known to be disease-related, while in the eight human signaling pathways only 25.5% (165 of 646) and among all human proteins listed by Ensembl only 20% (3929 of 19 534). For both comparisons P < 0.001.

3.4 Cross-species comparison of cross-talks

Next, we focused on how the complexity of intracellular signaling increases with a growing complexity of the organisms. In C.elegans only six of the eight curated pathways are active, and the Notch pathway is isolated (Fig. 2A). In addition, the cross-talk network of the pathways—where nodes represent pathways and links represent cross-talks—is sparse. Between the six active pathways only 5 of...
After merging SignaLink with protein expression profiles from TGF, and cross-talks from WNT to TGF and from EGF/MAPK to Notch, we found that cross-talks between the pathways EGF/MAPK and Notch are the least affected. See text for details.

To map some of these possibilities, we investigated the dynamical activity of cross-talks. Cross-talks, similar to other PPIs, are not active permanently in all tissue types. We considered an interaction to be possibly active in a given tissue type, if both of the mRNAs of its participating proteins are expressed in that tissue. It is reasonable to assume (as a simple approximation) that proteins whose mRNAs are expressed could be active in the given tissue (compared to those that are not transcribed).

After merging SignaLink with protein expression profiles from human colorectal, muscle, skin, liver and cardiovascular tissues, we quantified the possible tissue-specific activity of cross-talks. In C. elegans, cross-talk is possible through receptors, mediators and transcription factors (Fig. 2D). In the other two species, all pathway sections can participate in cross-talk, except for the NHR and JAK/STAT pathways of Drosophila, where cross-talk occurs mostly at the transcriptional level and through mediators, respectively (Fig. 2E and F).

In addition to the number of active pathways and cross-talks, a further important indicator of signaling complexity is the number of cross-talks relative to the number of pathways, especially in worm: 16 of the total 30 possible cross-talk types are present (Fig. 2B). In humans (the most complex organism of the three), all eight curated signaling pathways are active and almost all of the 56 possible cross-talk types are possible (Fig. 2C). The ubiquity of cross-talks (all 28 pathway pairs can cross-talk) expands both the repertoire of possible phenotypes and the system-level responses to environmental and pathological changes.

The presence of cross-talks in many pathways and pathway sections is a sign of the efficient utilization of resources: expanding the functions of an already existing pathway protein is more efficient than evolving a novel protein (Bhattacharyya et al., 2006). Given the high number of signaling cross-talks, a large variety of specific and robust phenotypes may emerge (Taniguchi et al., 2006). However, the actual signaling responses are controlled mainly by scaffold proteins, feedback loops, kinetic insulation, and the spatial and temporal expression patterns of proteins (Behar et al., 2007; Bhattacharyya et al., 2006; Freeman, 2000; Kholodenko, 2006).

To map some of these possibilities, we investigated the dynamical activity of signaling cross-talks in humans where cross-talk was found ubiquitous.

### 3.5 Tissue- and disease-specific activity of cross-talks

Cross-talks, similar to other PPIs, are not active permanently in all tissue types. We considered an interaction to be possibly active in a given tissue type, if both of the mRNAs of its participating proteins are expressed in that tissue. It is reasonable to assume (as a simple approximation) that proteins whose mRNAs are expressed could be active in the given tissue (compared to those that are not transcribed).

Average 26% and 48% of their proteins are expressed in the selected healthy tissue types (Fig. 3A).

Cancers are often viewed as systems diseases (Horberg et al., 2006). In cancer cells large-scale modifications of signaling pathways, especially changes of cross-talks (Stelling et al., 2004), are prevalent. Accordingly, detecting which proteins (cross-talks) are differentially expressed (active) in a carcinoma tissue may point out key causes of the given tumour and can help the identification of novel, systems-based drug targets (Korcsmaros et al., 2007; Tortora et al., 2004). To do this, we merged the network of eight human signaling pathways with protein expression data from human liver carcinomas. We considered a signaling interaction to be altered in these two liver carcinomas, if, compared to healthy liver tissues, at
According to a recent study (Cusick et al., 2007), we analyzed the drug target relevance of human signaling proteins by examining four key properties: disease-relatedness, localization in the plasma membrane (Gao et al., 2008; Yildirim et al., 2007), enzymatic functions and kinase domain content (Fabbo et al., 2002). To identify the most promising drug target candidates from the 327 proteins selected in the previous paragraph, we first listed those nine that have all four key properties, i.e. disease-related enzymes (with a kinase domain) localized in the membrane (Fig. 4). Out of these nine proteins five (CASK, EGFR, ErbB2, IGF1R, and INSR) are currently used as drug targets, while the other four (IRAK1, MAP3K13/LZK/MLK, ROR2 and TGFBR1) are not. ROR2 is essential during bone formation (Liu et al., 2007) and the other three proteins are inflammation-related factors (Klein et al., 2001). Interestingly, ROR2 was recently suggested as a therapeutic target for osteosarcoma based on expression analysis and siRNA treatment (Morikawa et al., 2009). As for IRAK1, TGFBR1 and MLK, anti-inflammatory drugs (imiquimod, dexamethasone and L-arginine, respectively) frequently affect their pathways, but without sufficient specificity (Klein et al., 2001). These findings support our predictions that these four human signaling proteins are promising novel drug targets.

4 DISCUSSION

4.1 Advantages and limitations of SignaLink
According to a recent study (Cusick et al., 2009), manual curation projects: (i) inherit the selection biases of the curated experiments; (ii) often lack the specific goals clearly defining the curation criteria; and (iii) it is usually difficult to estimate their completeness and reliability. Note, however, that many signaling proteins function outside the cell (ligands), in membrane-bound positions (receptors), or in the nucleus (transcription factors), and that proteins from these compartments are underrepresented in current HTP data. Thus, for the purpose of identifying signaling proteins and interactions, HTP techniques seem to be less adequate than those manual curation projects that have clear goals. In the case of SignaLink, the precisely defined curation process combines original research articles and reviews, thus, both experimental evidence and its critical discussion by specialists are included.

Applying a biochemically based, well-documented and clear pathway definition is central to SignaLink. For example, the EGF/MAPK pathway in SignaLink 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, EKR and ASK). An important consequence of precise pathway definitions is the reduced number of examined pathways. We suggest that appropriate and precise grouping, avoiding artificial pathway constructs, may be a better indicator of the goodness of the resource than merely the large number of pathways.

3.6 Potential role of cross-talking proteins in drug target discovery
Signaling proteins are overrepresented among human disease genes (Sakharkar et al., 2007) and are intensively studied as potential drug target candidates (Chaudihari and Chant, 2005), often with network-based methods (Berger and Iyengar, 2009). Among the 62 human multi-pathway proteins, 21 (33.8%) are known drug targets compared to 15% (94 of 646) in all 8 pathways and 8.2% (1610 of 19,534) among all human proteins. This implies that the remaining set of 41 human multi-pathway proteins may also be relevant for drug target discovery. In summary, the following two sets, altogether 327 proteins in our current study, are likely to be enriched with possible drug targets: (i) human multi-pathway proteins and (ii) proteins participating in cancer-related cross-talks.

In summary, the largest signaling pathway, EGF/MAPK, is frequently used in the selected tissue types and significantly changed in the two liver carcinomas. JAK/STAT is also strongly used, but less frequently used in the selected tissue types and significantly changed in the two investigated liver carcinomas. A third example is Notch, which is neither strongly used nor strongly modified. Thus, even though cross-talks are possible between all pairs of investigated human signaling pathways, the possible activity of these pathways in healthy tissues and the modifications of their cross-talks in the analyzed cancer types are highly diverse. See the Supplementary Material for additional literature support.

Fig. 4. Analysis of drug target candidate proteins listed with SignaLink. (A) The numbers of membrane proteins (m), enzymes (e), proteins with kinase domains (k) and disease-related proteins (d) among the three groups of proteins listed in SignaLink as possible novel drug targets (control: all SignaLink proteins). (B) Proteins with a fixed number of key properties out of the listed four (m, e, k and d). For each group the number of drug target proteins is shown separately. The names of the most promising candidates (have all four properties, not yet targeted) are listed.
Table 2. Comparison of database content for human pathways between three manually curated databases and SignaLink

<table>
<thead>
<tr>
<th>Pathways</th>
<th>KEGG</th>
<th>Reactome</th>
<th>NetPath</th>
<th>SignaLink</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proteins</td>
<td>429</td>
<td>65b (348)</td>
<td>687 (669)</td>
<td>96f (900)</td>
</tr>
<tr>
<td>Interactions</td>
<td>1502b (900)</td>
<td>362 (355)</td>
<td>457 (701)</td>
<td>646f</td>
</tr>
<tr>
<td>Cross-talks</td>
<td>225b (228)</td>
<td>79b (171)</td>
<td></td>
<td>300f</td>
</tr>
<tr>
<td>Number of publications</td>
<td>73</td>
<td>166</td>
<td>351</td>
<td>941</td>
</tr>
</tbody>
</table>

In each pairwise comparison, we compared only the pathways curated in both databases. For the same pathways in SignaLink, EGFR, EGFR1, and MAPK are compared to the EGFR/MAPK pathway of SignaLink.

Of the constantly increasing number of signaling databases (Bader et al., 2006) many are proprietary, list fewer than 200 molecules, or only selected types of pathway components, e.g. protein kinases. Even among the few databases passing these criteria (free for academic use, more than 200 molecules, all pathway component types included) there are currently no gold standards compiled with similar goals and methods as SignaLink. It is, therefore, important to compare both the curation protocols and the actual data of several available databases before selecting one of them for a particular analysis. In Table 2, we compare three widely used pathway databases—KEGG, Reactome, and NetPath—and SignaLink (see the Supplementary Material for details). In each pairwise comparison, we used the pathways available in both databases.

According to Table 2 and the Supplementary Material, SignaLink has the following advantages compared to the three analyzed databases: (i) precisely defined and documented curation protocol; (ii) highest numbers of signaling proteins and interactions in the curated signaling pathways; (iii) highest numbers of cross-talks and multi-pathway proteins; (iv) largest protein overlap with the other databases; (v) a higher than average number of publications used per pathway; (vi) minimal usage of protein isoform names; (vii) no binary interactions inferred from the membership of two proteins in the same complex; (viii) low number of proteins from UniProt/TrEMBL (i.e., few unverified proteins). SignaLink was compiled based on pathway reviews and primary research articles.

Thus, the high numbers of signaling proteins (including multi-pathway proteins) and cross-talks are likely to be dominated by true positives, indicating higher precision and coverage.

Despite the care we have taken in creating SignaLink, it does have limitations, e.g., SignaLink does not contain all signaling proteins. Only those signaling proteins have been included that have an experimentally verified function in the selected eight major pathways or a directed interaction with at least one of their proteins. Several groups of proteins were fully excluded from SignaLink. These groups, together with our detailed reasons for excluding them, are listed in the Supplementary Material. The compilation of SignaLink was based on published review and research papers. Note that the curation of the current version of SignaLink was closed in May 2008. Naturally, more pathways (defined by the same evolutionary and biochemical rules) and proteins can be added in a future version. We plan to update SignaLink every July (starting 2011). The next update will include recent high-confidence HTP data. Overall, based on the comparison we presented in this subsection, we expect that the limitations of SignaLink are small compared to the improvements it can provide.

4.2 Current applications

The primary goal of SignaLink is to provide maps of global pathway communication in three metazoans (C.elegans, D.melanogaster and H.sapiens) with well-documented, uniform manual curation. Interactions from all healthy tissue types were included into SignaLink, and expression data from selected tissues were used for dynamic analysis. We found that the pathways EGFR/MAPK, JAK/STAT and Notch are clear examples for three distinct types of behavior: (i) high expression in normal tissue types and strong changes in cancer; (ii) high expression, but small changes; and (iii) low expression and small changes. See the Supplementary Material for additional literature support.

Network analyses—combined with system-level resources—can contribute to modern drug target discovery, e.g. to polypharmacology and multi-target drug selection (Hopkins, 2008; Korcsmaros et al., 2007). With SignaLink, one can prioritize novel drug target candidates by examining the list of: (i) multi-pathway proteins and (ii) proteins participating in cancer-related cross-talks. Some of these proteins could be specific and proper targets, some of them could be too central and aspecific. After listing the properties of these proteins relevant for drug target selection, we suggested four novel drug target candidates. One of them, ROR2, was recently proposed as a novel chemotherapeutic target, while the other three are known to be non-specifically affected by anti-inflammatory drugs. A broader list contains 35 additional proteins with a lower confidence, which may be again filtered with additional criteria, e.g. phosphatase activity.
In biomedical experimental work, the functions of a few selected proteins are often modified. These changes can unexpectedly perturb signaling pathways and non-specifically affect cellular processes. Based on several data sources, including SignaLink, we have launched a web server, PathwayLinker, to aid experimental work by linking the queried proteins to signaling pathways through physical and/or genetic interactions (Farkas et al., submitted for publication).

4.3 Future applications

As their name suggests, targeting multi-pathway proteins may not be selective. However, selectivity is a key property in pharmacology, thus, analyses of multi-pathway proteins could support drug target discovery. Based on SignaLink, we suggest single gene knock-out experiments or RNA silencing of individual cross-talking proteins to help understand the functional selectivity of signaling pathways and to reduce the redundancies that are assumed to make many currently used drugs less efficient (Tortora et al., 2004; Urban et al., 2007).

For the mathematical modeling of the dynamical behavior of biomolecular pathways the precise, high-coverge reconstruction of the static network structure of these pathways is crucial (Papin et al., 2005). In the case of signaling pathways, the manual curation of the existing literature can be efficient for assembling large-scale interaction maps. SignaLink datasets have a simple and uniform structure and are available in several formats for all eight pathways and three species at http://SignaLink.org. This allows one to easily merge SignaLink with stoichiometric and expression data from, e.g. perturbation analyses (Papin et al., 2005). The static network provided by SignaLink can serve as a backbone for both numerical and differential equation-based models and—due to SignaLink’s focus on cross-talks—for decoupling the rules of cooperation between pathways (Borisov et al., 2009; Wang et al., 2009). Finally, perturbation analyses that cannot be carried out with HTP PPI networks (they list undirected PPIs), may be manageable with SignaLink, because it contains the directions of interactions.

A central goal of synthetic biology is to engineer cells that can carry out novel tasks (Bhattacharyya et al., 2006; Friedland et al., 2009). One way to achieve this goal is to rewire signaling circuits by modifying scaffold proteins and other biomolecules or by changing feed-back loops. In these studies, detailed high-quality maps of intracellular signaling are essential, especially the positions of cross-talks and multi-pathway proteins.

Comparative evolutionary studies usually focus on conserved and altered mechanisms underlying the observed differences in body plans. Among metazoans, these differences are largely due to changes in the complexity of regulation rather than different gene numbers (Levine and Tjian, 2003; Szathmary et al., 2003). Two decisive regulatory networks in this case are transcriptional used drugs less efficient (Tortora et al., 2004; Urban et al., 2007).

5 CONCLUSIONS

Contrary to earlier views, signaling pathways are now understood as interlinked (not linear) routes heavily interlinked by cross-talks. This major paradigm shift necessitates novel, systems-based approaches, e.g. new techniques for curation and modeling. Attempting to meet the novel curation requirements, we have compiled a signaling pathway resource, SignaLink. It allows the systematic comparison of pathways and their cross-talks. After finding that any two of the eight selected human signaling pathways may cross-talk, we quantified the possible activity patterns of these cross-talks in healthy and cancerous human tissues. Large-scale mapping of pathways and the activity patterns of their cross-talks revealed multi-pathway and cancer-related cross-talking proteins that can be relevant for drug target discovery.

ACKNOWLEDGEMENTS

We thank B. Papp, F. Jordan, and L. Zsakai for comments; P. Csernus for proofreading; T. Viczek for discussions; G. Szuroni, R. Palotai, and P. Pollner for technical help. The authors are grateful to the anonymous referees for their comments and suggestions.

Funding: EU 6th Framework Programme (FP6-518230); Hungarian National Science Fund (OTKA K69105, K75334, K68669); Hungarian National Office for Research and Technology (CellCom RET); EUE Fellowship (to I.L.F.).

Conflict of Interest: none declared.

REFERENCES


Downloaded from http://bioinformatics.oxfordjournals.org/ on August 24, 2010