KeyPathwayMinerWeb: online multi-omics network enrichment

Markus List[1,2,3,4,5,†], **Nicolas Alcaraz**[1,3,†], **Martin Dissing-Hansen**[1], **Henrik J. Ditzel**[2,3,6], **Jan Mollenhauer**[2,3] and **Jan Baumbach**[1,5,*]

[1]Department of Mathematics and Computer Science, University of Southern Denmark, 5230 Odense, Denmark, [2]Lundbeckfonden Center of Excellence in Nanomedicine NanoCAN, University of Southern Denmark, 5000 Odense, Denmark, [3]Institute of Molecular Medicine, University of Southern Denmark, 5000 Odense, Denmark, [4]Institute of Clinical Research, University of Southern Denmark, 5000 Odense, Denmark, [5]Max Planck Institute for Informatics, 66123 Saarbrücken, Germany and [6]Department of Oncology, Odense University Hospital, 5000 Odense, Denmark

ABSTRACT

We present KeyPathwayMinerWeb, the first online platform for *de novo* pathway enrichment analysis directly in the browser. Given a biological interaction network (e.g. protein–protein interactions) and a series of molecular profiles derived from one or multiple OMICS studies (gene expression, for instance), KeyPathwayMiner extracts connected subnetworks containing a high number of active or differentially regulated genes (proteins, metabolites) in the molecular profiles. The web interface at (http://keypathwayminer.compbio.sdu.dk) implements all core functionalities of the KeyPathwayMiner tool set such as data integration, input of background knowledge, batch runs for parameter optimization and visualization of extracted pathways. In addition to an intuitive web interface, we also implemented a RESTful API that now enables other online developers to integrate network enrichment as a web service into their own platforms.

INTRODUCTION

Interaction networks are at the core of systems biology research ever since it became possible to detect protein–protein interactions. Since then, more advanced detection methods have been developed and with them, biological interaction networks have been growing (1) from a few thousand interactions as in the Human Protein Reference Database (2) to several hundred thousand interactions as in BioGrid (3), IntAct (4) or I2D (5). Together with the continuous growth of molecular interaction information, research efforts in systems biology have been directed toward meaningful ways of integrating biological networks with molecular profiles (6).

Exploiting current interaction databases has led to the development of pathway-level enrichment methods for standard downstream analyses in biological and biomedical settings. In their simplest form, classical pathway enrichment approaches attempt to aggregate the individual measurements of genes (or their products) in a pathway to produce a single score representing the pathway's level of activity or deregulation. However, these methods rely on a pre-defined list of pathways of known biological processes that play a role in normal or diseased cell function. This may bias the search towards known pathways and overlook unknown, yet important functional modules that may be just a small part of or completely independent from any of the pathways available.

To overcome this limitation, so-called *de novo* network enrichment approaches have become increasingly popular. A wide range of methods have spawned inspired by the pioneering work of Ideker *et al.* (7). Although existing approaches differ in many relevant aspects (optimization criteria, algorithmic implementation, scoring function, etc.), they all aim for extracting connected subnetworks from a larger interaction network. These are significantly enriched with active, i.e. deregulated, biological entities (genes, proteins, metabolites). Popular methods are JActiveModules (7), BioNet (8), DEGAS (9), GiGa (10), HotNet (11) and KeyPathwayMiner. Drug target identification, functional annotation and biomarker discovery are some of the various applications *de novo* network enrichment methods offer to the biomedical community.

We have previously developed and extended KeyPathwayMiner, a set of *de novo* network enrichment methods for extracting condition-specific pathways from single or multiple OMICS datasets in a flexible and intuitive manner (12–14). Note that KeyPathwayMinerWeb can handle different

*To whom correspondence should be addressed. Tel: +45 6550 2387; Fax: + 45 6550 2373; Email: jan.baumbach@imada.sdu.dk
†These authors contributed equally to the paper as first authors.

and multiple OMICS data types. However, to improve readability, in the remainder of this article we assume a given case/control gene expression dataset and use corresponding nomenclature, although KeyPathwayMinerWeb would work with any OMICS dataset as long as the IDs of the expression study match the IDs in the utilized (or uploaded) network.
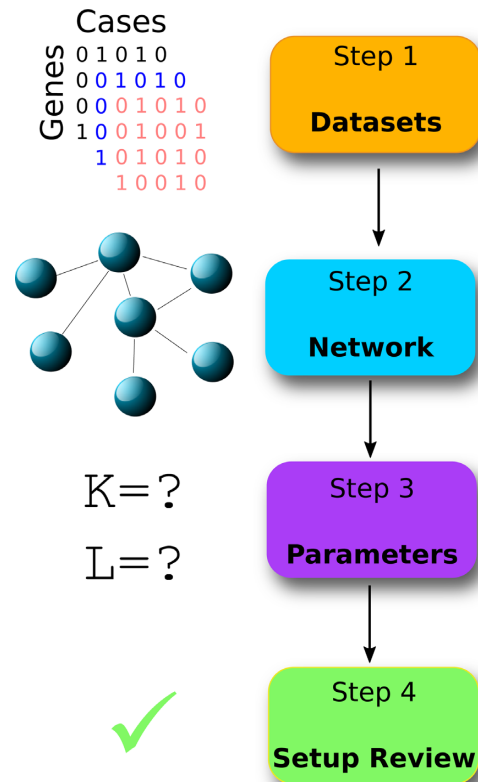
Similar to other *de novo* network enrichment tools, KeyPathwayMiner is integrated into the network visualization and analysis framework Cytoscape (15). Alternatively, *de novo* network enrichment is available in scripting languages such as R (8). However, the user experience in Cytoscape as well as in scripting languages suffers from a steep learning curve. This limits the use of *de novo* network enrichment tools for biomedical researchers, which rely on user-friendly and intuitive tools. Preferably such tools should be accessible without technical barriers. Here, web applications are superior to desktop applications, since they do not have any local dependencies and do not have to be installed. To our knowledge, however, no *de novo* network enrichment tool is available as a web application yet.

This motivated us to develop KeyPathwayMinerWeb, a web frontend for the KeyPathwayMiner software library, providing a responsive and interactive user interface as well as a RESTful API allowing other developers to integrate *de novo* network enrichment as a web service.

KEYPATHWAYMINER

In KeyPathwayMiner, two different approaches for extracting subnetworks that are enriched for active/deregulated genes have been implemented. For the INES (Individual Node Exceptions) approach, two parameters are required. A gene is considered foreground, if it is active, e.g. differentially expressed, in all but $L\%$ of the samples (cases, cell lines or patients). In addition, a parameter K adjusts for the number of inactive genes (exceptions, background) that are allowed in a solution. Once K and L have been selected, KeyPathwayMiner then proceeds to extract all maximal subnetworks containing at most K (exception) nodes with no more than $L\%$ inactive cases. In some instances, exception nodes may be part of, but are not central to the pathway (e.g. removing them would maintain the pathway connected) . Hence, we implemented a new optional post-processing step to remove Border Exception Nodes (BENs).

The parameter K tends to allow KeyPathwayMiner to select hub nodes to combine small solutions into large connected ones. Since this behavior is not always desired, KeyPathwayMiner also implements a second strategy called GLONE (Global Node Exceptions). Here, the parameter K is omitted in favor of a global view on the parameter L. A subnetwork is considered foreground (active), if the total sum of inactive cases across all genes is smaller than or equal to L. Here we also aim at extracting all maximal connected sub-networks containing at most L inactive cases in total. This strategy is less prone to selecting solutions containing hub nodes, but it is computationally more expensive. For details on the implementation as well es extensive evaluations and application examples of the KeyPathwayMiner methodology we refer to (12–14,16).

Figure 1. The user is guided through the setup of a *de novo* network enrichment analysis in KeyPathwayMinerWeb in four steps. The only prerequisite is the creation of suitable input matrices for the multi-OMICS datasets. Tutorials, sample data files and a screencast are available at the web site.

INPUT

KeyPathwayMinerWeb offers *de novo* network enrichment using the INES and GLONE strategies introduced above. The user is guided through the setup of an analysis in four steps (Figure 1):

Step 1: setup datasets

Select datasets. In the first step, the user is expected to select or upload one or several datasets. For each case/sample, a particular gene can be reported as deregulated (active or not). This information is encoded in a data matrix, where columns correspond to cases (without column names) and rows correspond to genes (with the gene IDs in the first column). Each of the table entries encodes a '1' for active and '0' for inactive case-gene combinations. Columns need to be separated by a tab delimiter. Instead of such an indicator matrix, users can also upload an arbitrary numerical matrix of P-values or fold changes, for instance. In the latter case, the user can select a threshold for each dataset and an indicator matrix is generated on the fly. Note that P-values in a typical statistical analyses are derived from a comparison between a case and a control group. However, this leads to an aggregated P-value that indicates if the observed difference was overall significant but it does not reflect the fact that in biomedical data, nodes (genes, proteins, metabolites, etc.) of interest are typically deregu-

lated in many of the cases with a certain number of exceptions. The key advantage of KeyPathywayMiner is that it enables its users to control the number of these case exceptions. Consequently, a differential analysis should ideally not be focused on a group-wise comparison but produce a *P*-value for each case. This can be achieved, for instance, by comparing matched case/control samples or by comparing each case to a distribution of values derived from a control group. To demonstrate the makeup of the expected input data, KeyPathwayMinerWeb includes a number of demo datasets, such as indicator matrices of gene expression (up- and downregulated) and DNA methylation in colon cancer (17) as well as p-values for differential gene expression in Huntington's disease (HD) (12).

Logically connect experiments. This field specifies how KeyPathwayMinerWeb deals with multiple datasets. Selecting 'AND' results in a gene-case combination to be considered as active if it is active in all of the selected datasets. In contrast, selecting 'OR' results in a gene-case combination being considered as active if it is active in any of the selected datasets.

Positive and negative nodes. Sometimes, users have previous knowledge that they may want to include in the analysis, for instance, by preferring specific genes of interest in a solution or by punishing solutions with unfavorable genes. To facilitate this, users can add genes either to the positive (always considered active) or the negative (always considered inactive) list. KeyPathwayMinerWeb will favor subnetworks with many positive nodes and few negative ones.

Step 2: setup network

In addition to selecting one or more datasets, a network can either be uploaded or selected from a list of provided default networks, including HPRD (2), BioGrid (3) or I2D (5), for instance. User provided network files need to follow the simple interaction file (SIF) format. The gene IDs need to match the gene IDs used in the OMICS datasets. Note that all default data and networks in KeyPathwayMinerWeb use Entrez gene IDs. We provide template files as well as a detailed description of the data format at the web site.

Step 3: setup parameters

In this step of the setup, the user can optionally enter a custom name for the analysis or stick to an automatically generated one. The job name will later be used to select and analyze the results. Next, one has to choose the parameters needed by KeyPathwayMiner, starting with the enrichment strategy (INES or GLONE) and the corresponding parameters *K* (node exceptions) and *L* (case exceptions in percent of total cases). The user may compare the results for different parameters conveniently by defining a range for *K* and *L*. Not all of the genes in a network are necessarily represented in all or any of the datasets. The user can choose whether these unmapped genes should be added to the positive list (and thus extend and connect extracted subnetworks) or to the negative list (where they will be ignored). Finally, the user can select how many enriched subnetworks KeyPathwayMiner extracts with an upper limit of 20.

Step 4: setup review

In the final step, the user can review the setup in a summary before triggering the KeyPathwayMiner analysis with pressing the start button. The user will then automatically be directed to the results tab.

OUTPUT

Job overview

The status of all queued, active and completed runs of a user session is shown in the results tab. For queued jobs, the current position in the queue is shown. For jobs that are currently active, a progress bar is shown. Once a job is completed, a 'See results' button appears.

Result view

At the top of the result page (Figure 2A), the user may first select a specific *L* (and *K*, depending on the selected enrichment strategy) in case a parameter range was chosen. Optionally, users can choose to substitute node labels with a different identifier type (Figure 2B). This is a convenience feature that allows, for instance, to show familiar gene symbols instead of cryptic gene IDs.
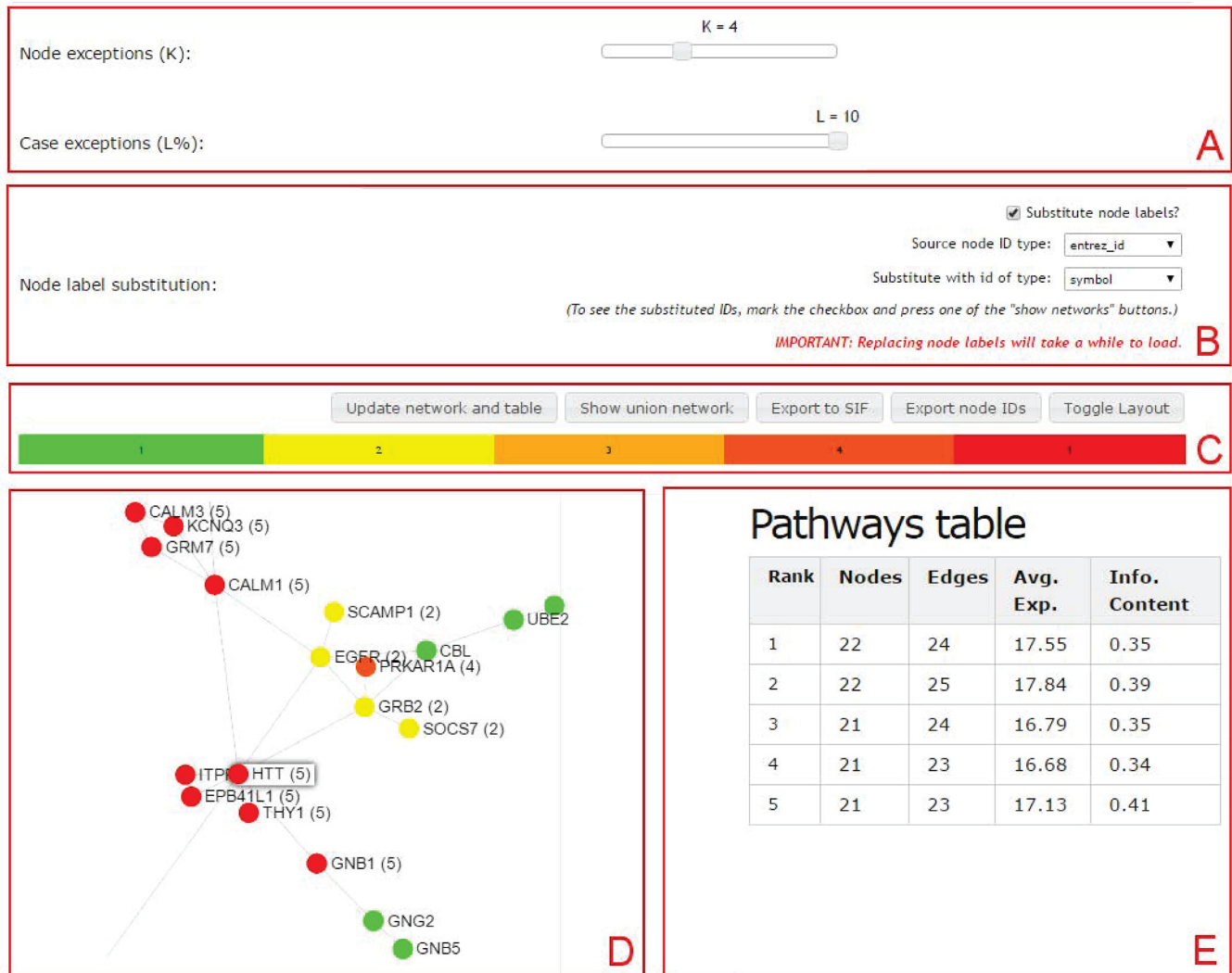
The central element of the result view is an interactive graph panel (Figure 2D). To avoid cluttering in large solutions, node labels are omitted and only shown when the mouse cursor is moved over a specific node or for all visible nodes when zooming in until cluttering is minimal. Clicking on a node will open a new tab where additional information about a particular gene is shown via the NCBI website. The graph panel is accompanied by the pathways table (Figure 2E) that summarizes the extracted subnetworks and reports the number of nodes and edges, the average expression, which corresponds to the average number of active cases, as well as the average information content. Clicking on an entry of the pathways table will update the graph plot.

Additional options are available through buttons shown on top of the graph plot (Figure 2C). This includes the possibility to update the graph, for instance after changing the node label substitution settings or to export the currently selected solution as a list of nodes or as a SIF file. Finally, the user can also render a union graph, in which all extracted subnetworks are merged to build a consensus network. In the union graph, the node color can be interpreted with the help of a color legend, which indicates how often a particular gene is part of a reported solution.

IMPLEMENTATION

Web application framework

We chose a Java web application framework for the development of KeyPathwayMinerWeb to exploit the fact that the KeyPathwayMiner algorithms have been implemented in Java. Several factors motivated us to choose the Grails framework as a foundation. First of all, Grails builds on top of the Spring framework, which is an enterprise standard for the development of Java web applications. Moreover, Grails embraces concepts that reduce the development effort and increase maintainability, such as convention over

Figure 2. Example output of a typical KeyPathwayMinerWeb analysis using differential gene expression data from 38 Huntington disease patients and 32 control subjects. (**A**) KeyPathwayMinerWeb allows users to explore several parameter combinations. Sliders allow changing the parameters conveniently. (**B**) Node labels can be substituted with alternative identifiers, such as gene symbols. (**C**) Additional options for data export and for interacting with the results are available, including a union graph feature. (**D**) An interactive representation of the selected subnetwork or the union graph. (**E**) A summary table showing the solutions, i.e. the enriched subnetworks. Clicking on a table row will update the graph shown in (D).

configuration or separation of concerns. Finally, Grails reduces the complexity of data persistence with an object relational modeling technique, which encapsulates all database interactions and models them through simple Java domain classes.

User authentication

For each session, KeyPathwayMinerWeb creates a unique session ID, which enables users to perform analyses anonymously. However, in order to manage larger number of analyses and to prevent the results from being deleted, users can optionally create a user account. Secure user management and authentication is implemented via the Grails SpringSecurity plug-in.

Queuing

One crucial aspect that was not covered by the Grails plug-ins ecosystem was a queuing system that would allow us to provide a stable and responsive user experience while handling many user requests in parallel. We thus implemented a queuing mechanism for KeyPathwayMinerWeb that provides users with the possibility to monitor for each of the submitted jobs the position in the queue. When a job is first in the queue, a KeyPathwayMiner thread is created and its progress is reported back to the user via asynchronous frontend updates. Consequently, the user will be able to follow the progress exactly as if the job was executed locally in the Cytoscape app.

Network visualization

For the visualization of the extracted subnetworks, we used sigmajs (http://sigmajs.org/, last access 5 April 2016), a state

of the art Javascript library for interactive graph visualization.

Identifier conversion

KeyPathwayMinerWeb requires the dataset and the network to use the same identifier. However, it is possible to convert the IDs of the extracted subnetworks, to show, for instance, better interpretable gene symbols instead of Entrez gene IDs. This optional mapping step is realized through integrating the web service of the HUGO Gene Nomenclature Committee (http://www.genenames. org/help/rest-web-service-help, last access 5 April 2016).

RESTful API

The functionality of KeyPathwayMinerWeb can be integrated into third party applications through a RESTful API. The API allows jobs to be submitted either in a blocking (synchronous) or non-blocking (asynchronous) way, depending on the needs of the developer. A list of integrated default networks is available, allowing developers to offer their users to choose one of them. Alternatively, external applications may submit their own network in SIF format. All files need to be base64 encoded prior to submission. In order to optimally support asynchronous job execution, there are methods available for querying the status and progress of a job as well as for obtaining the final results after the job is completed. The API is documented at http://keypathwayminer.compbio.sdu.dk/ documentation/rest and includes various usage examples in the R scripting language.

BIOLOGICAL USE CASE

We demonstrate the usability of KeyPathwayMinerWeb by analyzing a whole-genome gene expression dataset on HD (18) together with the human interactome (19). The expression dataset comprises samples taken from caudate nucleus region of the brain of 38 HD patients compared to 32 patients in a control group. One-tailed P-values were computed for every gene and HD sample compared to a normal distribution of the corresponding gene in the control group. Note that HD is a Mendelian disease, i.e. a certain mutation in the Huntingtin (HTT) gene triggers the disease. However, in the dataset HTT is not significantly differentially expressed in ca. 40% of the patients, such that a gene expression dataset analysis alone would not lead to the identification of this gene as significant. Here, we now aim to use KeyPathwayMinerWeb to identify downregulated pathways. To ease reproducibility we have integrated both, the HD expression data (as P-values) and the corresponding protein–protein network with the web platform already. One would start a new 'run' by selecting the down-regulated Huntington gene expression dataset as well as a P-value threshold (here: 0.05). Afterward, one needs to select a matching network resource (here: the Ulitsky homo sapiens interactome with Entrez IDs) as well as the KeyPathwayMiner parameters (here: INES algorithm, no BENs, $k = 4$, L = 10%). KeyPathwayMinerWeb then extracts the five largest connected pathways where all genes but four are downregulated in all

HD patients but at most 10%. The result is shown in Figure 2D as union graph representation across the five best solutions (33 nodes/genes). The HTT gene is found in the center as exception node that is highly connected to differentially expressed genes in all five solutions. Note that 9 of the 33 (27%) genes (incl. HTT) in the reported five best solutions are known to be related to HD (19).

DISCUSSION AND CONCLUSION

It is widely acknowledged that a narrow view on a single data type such as gene expression can only offer limited insights into the complex mechanisms of bio-molecular systems and their perturbation in complex diseases. A current trend in systems biology is therefore the integration of multiple OMICS datasets. Molecular interaction networks are ideally suited to facilitate this integration, since they allow for the unbiased identification of subnetworks or pathways affected by an experiment through *de novo* network enrichment.

However, existing methods are not easily accessible for biomedical researchers, which motivated us to develop the first web application that supports this type of analysis. An emphasis of KeyPathwayMinerWeb is to be user-friendly. The user is guided step by step through setting up a *de novo* network enrichment analysis. An online video screencast (available at the KeyPathwayMinerWeb front page), as well as a selection of default datasets and networks allows users to quickly familiarize themselves with the tool, equipping them with the necessary knowledge to analyze their own data successfully.

While the analysis with preprocessed datasets is relatively straight forward, we observe that the generation of suitable indicator matrices from experimental data remains quite challenging. Although this is a direct consequence of the flexibility of KeyPathwayMiner with regards to the support of arbitrary data types, we sought a way to overcome this limitation. To this end, we implemented support for non-indicator matrix input (e.g. P-values or fold changes), such that the user can provide a given cutoff for (differential) activity and produce the indicator matrix on the fly. Furthermore, a RESTful API allows developers of data analysis software to integrate *de novo* network enrichment as part of their downstream analysis. Developers of such an application can subsequently direct their users to the result page of KeyPathwayMinerWeb or query the results of the analysis for further use. One use case of the RESTful API is found in HiTSeekR, a platform for processing of high-throughput screening data from raw microtiter plate readout data down to interpretable systems biology results (http://hitseekr.compbio.sdu.dk, manuscript under review).

In addition to the non-indicator matrix input support and RESTful API, the new KeyPathwayMinerWeb release now also supports the following new functionalities not available in KeyPathwayMiner 4.0: (i) a gene ID to gene name mapping on the resulting pathways, (ii) a union graph/network as a consensus/summary representation of the results, (iii) a functionality to remove so-called 'BENs' (Boundary Exception Nodes).

KeyPathwayMinerWeb is significantly different from existing web resources. DAVID (20) and PANTHER (21), for

instance, are very popular services that essentially perform functional annotations for a set of input genes using, for instance, KEGG (22) pathway mapping and GO-term enrichment. GeneMANIA, in contrast, allows to connect a given set of input genes (while ignoring the experimental data itself) by using network information (23). ConsensusPathDB (24) and NetworkTrail (25) report for a given list of input genes their enrichments in pathway sets, such as KEGG pathways. All those platforms, in contrast to KeyPathwayMinerWeb, feature non-*de novo* methods for the extraction of functional network modules.

In spite of its potential for delivering valuable biomedical insights into the results of increasingly complex experimental setups, the practical application of *de novo* network enrichment in research is still limited. However, it is quite telling that the most successful systems biology methods developed so far (at least in terms of total number of citations), only became widely used in biomedical research after user-friendly applications became available. We expect that the lack of such an intuitive tool for *de novo* network enrichment currently still hinders much wider use of this kind of methodology. Consequently, we expect that KeyPathwayMinerWeb has the potential to permanently add *de novo* network enrichment to the stack of methods routinely used in systems biology analyses. To date, network enrichment is mainly applied to gene expression datasets, with few exceptions. With intuitive user interfaces, it will prove useful for other OMICS data types as well, ranging from proteomics (26) to metabolomics (27) and cancer pathway mutation screening (28), to name just a few examples.

FUNDING

REFERENCES

1. Röttger,R., Rückert,U., Taubert,J. and Baumbach,J. (2012) How little do we actually know? On the size of gene regulatory networks. *IEEE/ACM Trans. Comput. Biol. Bioinform.*, **9**, 1293–1300.
2. Prasad,T. K., Goel,R., Kandasamy,K., Keerthikumar,S., Kumar,S., Mathivanan,S., Telikicherla,D., Raju,R., Shafreen,B., Venugopal,A. *et al.* (2009) Human Protein Reference Database–2009 update. *Nucleic Acids Res.*, **37**, D767–D772.
3. Chatr-aryamontri,A., Breitkreutz,B.-J., Oughtred,R., Boucher,L., Heinicke,S., Chen,D., Stark,C., Breitkreutz,A., Kolas,N., O'Donnell,L. *et al.* (2015) The BioGRID interaction database: 2015 update. *Nucleic Acids Res.*, **43**, D470–D478.
4. 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**, 358–363.
5. Brown,K.R. and Jurisica,I. (2007) Unequal evolutionary conservation of human protein interactions in interologous networks. *Genome Biol.*, **8**, R95.
6. Baumbach,J. and Apeltsin,L. (2008) Linking Cytoscape and the corynebacterial reference database CoryneRegNet. *BMC Genomics*, **9**, 184.
7. Ideker,T., Ozier,O., Schwikowski,B. and Siegel,A.F. (2002) Discovering regulatory and signalling circuits in molecular interaction networks. *Bioinformatics*, **18**(Suppl. 1), S233–S240.
8. Beisser,D., Klau,G.W., Dandekar,T., Müller,T. and Dittrich,M.T. (2010) BioNet: an R-Package for the functional analysis of biological networks. *Bioinformatics*, **26**, 1129–1130.
9. Ulitsky,I., Krishnamurthy,A., Karp,R.M. and Shamir,R. (2010) DEGAS: de novo discovery of dysregulated pathways in human diseases. *PLoS One*, **5**, e13367.
10. Breitling,R., Amtmann,A. and Herzyk,P. (2004) Graph-based iterative Group Analysis enhances microarray interpretation. *BMC Bioinformatics*, **5**, 100.
11. Vandin,F., Upfal,E. and Raphael,B.J. (2011) Algorithms for detecting significantly mutated pathways in cancer. *J. Comput. Biol.*, **18**, 507–522.
12. Alcaraz,N., Kücük,H., Weile,J., Wipat,A. and Baumbach,J. (2011) KeyPathwayMiner: detecting case-specific biological pathways using expression data. *Internet Math.*, **7**, 299–313.
13. Alcaraz,N., Friedrich,T., Kötzing,T., Krohmer,A., Müller,J., Pauling,J. and Baumbach,J. (2012) Efficient key pathway mining: combining networks and OMICS data. *Integr. Biol.*, **4**, 756–764.
14. Alcaraz,N., Pauling,J., Batra,R., Barbosa,E., Junge,A., Christensen,A., Azevedo,V., Ditzel,H.J. and Baumbach,J. (2014) KeyPathwayMiner 4.0: condition-specific pathway analysis by combining multiple omics studies and networks with Cytoscape. *BMC Syst. Biol.*, **8**, 99.
15. Shannon,P., Markiel,A., Ozier,O., Baliga,N.S., Wang,J.T., Ramage,D., Amin,N., Schwikowski,B. and Ideker,T. (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.*, **13**, 2498–2504.
16. Pauling,J.K., Christensen,A.G., Batra,R., Alcaraz,N., Barbosa,E., Larsen,M.R., Beck,H.C., Leth-Larsen,R., Azevedo,V., Ditzel,H.J. *et al.* (2014) Elucidation of epithelial-mesenchymal transition-related pathways in a triple-negative breast cancer cell line model by multi-omics interactome analysis. *Integr. Biol.*, **6**, 1058–1068.
17. Hinoue,T., Weisenberger,D.J., Lange,C.P.E., Shen,H., Byun,H.-M., Van Den Berg,D., Malik,S., Pan,F., Noushmehr,H. *et al.* (2012) Genome-scale analysis of aberrant DNA methylation in colorectal cancer. *Genome Res.*, **22**, 271–282.
18. Hodges,A., Strand,A.D., Aragaki,A.K., Kuhn,A., Sengstag,T., Hughes,G., Elliston,L.A., Hartog,C., Goldstein,D.R., Thu,D. *et al.* (2006) Regional and cellular gene expression changes in human Huntington's disease brain. *Hum. Mol. Genet.*, **15**, 965–977.
19. Ulitsky,I., Karp,R.M. and Shamir,R. (2008) Detecting disease-specific dysregulated pathways via analysis of clinical expression profiles. In: Vingron,M and Wong,L (eds). *Research in Computational Molecular Biology*. Springer, Berlin Heidelberg, pp. 347–359.
20. Dennis,G., Sherman,B.T., Hosack,D.A., Yang,J., Gao,W., Lane,H.C. and Lempicki,R.A. (2003) DAVID: database for annotation, visualization, and integrated discovery. *Genome Biol.*, **4**, P3.
21. Thomas,P.D., Campbell,M.J., Kejariwal,A., Mi,H., Karlak,B., Daverman,R., Diemer,K., Muruganujan,A. and Narechania,A. (2003) PANTHER: a library of protein families and subfamilies indexed by function. *Genome Res.*, **13**, 2129–2141.
22. 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.
23. Warde-Farley,D., Donaldson,S.L., Comes,O., Zuberi,K., Badrawi,R., Chao,P., Franz,M., Grouios,C., Kazi,F., Lopes,C.T. *et al.* (2010) The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. *Nucleic Acids Res.*, **38**(Suppl. 2), W214–W220.
24. Kamburov,A., Stelzl,U., Lehrach,H. and Herwig,R. (2013) The ConsensusPathDB interaction database: 2013 update. *Nucleic Acids Res.*, **41**, D793–D800.
25. Stöckel,D., Müller,O., Kehl,T., Gerasch,A., Backes,C., Rurainski,A., Keller,A., Kaufmann,M. and Lenhof,H.-P. (2013) NetworkTrail–a

web service for identifying and visualizing deregulated subnetworks. *Bioinformatics*, **29**, 1702–1703.

26. Ertle,J.M., Heider,D., Wichert,M., Keller,B., Kueper,R., Hilgard,P., Gerken,G. and Schlaak,J.F. (2013) A combination of α-fetoprotein and des-γ-carboxy prothrombin is superior in detection of hepatocellular carcinoma. *Digestion*, **87**, 121–131.

27. Baumbach,J., Bunkowski,A., Lange,S., Oberwahrenbrock,T., Kleinboelting,N., Rahmann,S. and Baumbach,J.I. (2007) IMS2—an integrated medical software system for early lung cancer detection using ion mobility spectrometry data of human breath. *J. Integr. Bioinform.*, **4**, 1–12.

28. Leiserson,M. D.M., Vandin,F., Wu,H.-T., Dobson,J.R., Eldridge,J.V., Thomas,J.L., Papoutsaki,A., Kim,Y., Niu,B., McLellan,M. *et al.* (2014) Pan-cancer network analysis identifies combinations of rare somatic mutations across pathways and protein complexes. *Nat. Genet.*, **47**, 106–114.