techniques and applications available, the features that ViSP offers are of critical importance in the greater microscopy research community. ViSP is controlled with a simple graphical user interface and is compatible with Windows and Mac OS X. The software is freely available for academic use (source code is available upon signing a Material Transfer Agreement), and the latest versions can be downloaded at http://umr168.curie.fr/en/research-groups/locco/software/. The authors request acknowledgment of the use of ViSP in published works.

Note: Supplementary information is available in the online version of the paper (doi:10.1038/nmeth.2566)

ACKNOWLEDGMENTS

We thank B. Hajj (Institut Curie), J. Chen (Janelia Farm Research Campus), I. Izeddin (Ecole Normale Supérieure), J. Sillibourne (Commissariat à l'Energie Atomique et aux Energies Alternatives) and M. Bornens (Institut Curie) for providing super-resolution data sets used in the figures of this correspondence. We are grateful to K. Busch (Universität Osnabrück) for sharing the plasmid TOM-20-GFP. We also thank N. Clack, F. Mueller and X. Darzacq for discussions during the preparation of this work. This work was supported by funding from the Transcription Imaging Consortium at the Janelia Farm Research Campus, the Institut Curie International PhD Program and by the Tridimic grant from the program Recherches Partenariales et Innovation Biomédicale from Agence Nationale pour la Recherche.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Mohamed El Beheiry[1,2] & Maxime Dahan[1,2]

[1]Laboratoire Physico-Chimie Curie, Institut Curie, Centre National de la Recherche Scientifique Unité Mixte de Recherche 168, Paris, France. [2]Transcription Imaging Consortium, Howard Hughes Medical Institute, Janelia Farm Research Campus, Ashburn, Virginia, USA.
e-mail: mohamed.el-beheiry@curie.fr or maxime.dahan@curie.fr

1. Huang, B., Babcock, H. & Zhuang, X. *Cell* **143**, 1047–1058 (2010).
2. Kopek, B.G., Shtengel, G., Xu, C.S., Clayton, D.A. & Hess, H.F. *Proc. Natl. Acad. Sci. USA* **109**, 6136–6141 (2012).
3. Henriques, R. *et al. Nat. Methods* **7**, 339–340 (2010).
4. Wolter, S. *et al. Nat. Methods* **9**, 1040–1041 (2012).
5. Baddeley, D., Cannell, M.B. & Soeller, C. *Microsc. Microanal.* **16**, 64–72 (2010).
6. Gehlenborg, N. & Wong, B. *Nat. Methods* **9**, 851 (2012).

mentha: a resource for browsing integrated protein-interaction networks

To the Editor: Systems-level approaches require access to comprehensive genome-wide and proteome-wide databases. A comprehensive resource that archives all published protein-protein interactions (PPIs) is not available. In fact, primary PPI databases capture only a fraction of published data.

This dispersion of information has motivated projects such as the Agile Protein Interaction DataAnalyzer (APID), the Protein Interaction Network Analysis (PINA) platform, iRefWeb, Michigan Molecular Interactions (MiMI) and the Search Tool for the Retrieval of Interacting Genes/Proteins (STRING), which offer wider coverage of PPI information by integrating heterogeneously curated data. The difficulty of combining annotations from heterogeneous efforts, however, consistently hampers the integration of data extracted from databases that adopt different curation policies; one consequence of laborious integration procedures is that updates are infrequent.

Here we report mentha (http://mentha.uniroma2.it/), a PPI resource that takes advantage of the recent establishment of the International Molecular Exchange (IMEx)[1] consortium and the development of the Proteomics Standard Initiative Common Query Interface (PSICQUIC)[2] for automatic access to molecular-interaction databases. mentha integrates protein-interaction data curated by experts in compliance with IMEx curation policies, using the PSICQUIC protocol to implement an automatic procedure that, every week, without human intervention, aligns the integrated database with data regularly annotated by the primary databases (**Supplementary Methods**).

The scope and motivation behind mentha are different from those of databases such as STRING, which integrate information extracted with text mining and prediction methods. mentha favors precision over comprehensiveness, and it focuses on experimentally determined direct protein interactions (**Supplementary Note 1**). We note that the number of interactions and proteins archived in mentha is limited by the fact that it contains data annotated exclusively in primary PPI databases, without any inference.

In designing mentha we made the following choices: (i) to focus on experimentally demonstrated physical interactions, trying to avoid confusion between physical and genetic interactions and between experimental and inferred interactions; (ii) to maintain links to original articles and primary databases; and (iii) to preserve, as much as possible, the richness of the original annotation. We restrict the integration to databases that adopt the PSI-MI controlled vocabularies[3] and the IMEx curation policies. This choice, though it excludes the use of data-rich resources that have not yet adopted the IMEx standard, such as the Human Protein Reference Database, allows for higher data consistency. As a consequence, the integration procedure in mentha can make use of specific attributes assigned according to the common curation policy, such as "interaction type" and "interaction method," to assign a reliability score to each interaction, similarly to the Molecular Interactions scoring function[4]. The reliability score can be used to filter the PPI network of interest from

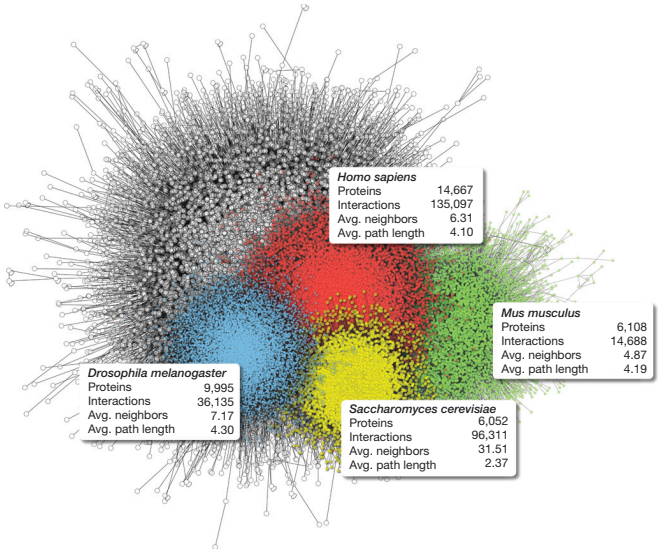


Figure 1 | mentha's interactomes. The gray graph illustrates mentha's "All" interactome. The colored graphs report the interactomes of *Homo sapiens* and three model organisms. The insets report the number of proteins, interactions and some topological characteristics. mentha offers graph analysis tools to extract subnetworks and paths, optionally identifying enzymatic interactions.

the intrinsic noise of PPI information (**Supplementary Note 2**). In fact, independent experimental approaches can yield contradictory PPI information, and inconsistent data curation, or curation errors, can result in inaccurate annotation[5]. Combining the evidence from different experimental approaches can increase the confidence in any specific binary interaction.

mentha archives PPI data for many species, including human; these data are updated weekly, and backups for past releases are available for download. mentha was designed as a workbench where the user can assemble and analyze collections of proteins and networks of interest (**Fig. 1** and **Supplementary Note 3**). mentha—the interactome browser—is accessible via a user-friendly website and via a RESTful Application Programming Interface. It also offers an interactive graphical application that can be embedded in web pages[6].

Note: Supplementary information is available in the online version of the paper (doi:10.1038/nmeth.2561).

ACKNOWLEDGMENTS

mentha is supported by the European Union Framework project 7 "Affinomics" grant and the "Oncodiet" grant from the Fondo per gli Investimenti della Ricerca di Base (FIRB) project.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Alberto Calderone[1], Luisa Castagnoli[1] & Gianni Cesareni[1,2]

[1]Department of Biology, University of Rome Tor Vergata, Rome, Italy. [2]Fondazione Santa Lucia Istituto di Ricovero e Cura a Carattere Scientifico (IRCCS), Rome, Italy. e-mail: sinnefa@gmail.com

1. Orchard, S. *et al. Nat. Methods* **9**, 345–350 (2012).
2. Aranda, B. *et al. Nat. Methods* **8**, 528–529 (2011).
3. Hermjakob, H. *et al. Nat. Biotechnol.* **22**, 177–183 (2004).
4. Licata, L. *et al. Nucleic Acids Res.* **40**, D857–D861 (2012).
5. Salwinski, L. *et al. Nat. Methods* **6**, 860–861 (2009).
6. Liberti, S. *et al. FEBS J.* **280**, 379–387 (2013).

Convergence of chromatin binding estimates in live cells

To the Editor: Many transcription factors exhibit highly dynamic interactions with chromatin as measured by different fluorescence light microscopy techniques in live-cell nuclei[1]. There is mounting evidence that this transient binding quantitatively influences the process of transcription from target genes[1]. As a result, much effort has been devoted over the past decade to quantify these binding interactions, namely, to determine the fraction of transcription factors bound and their average residence time on chromatin.

Early studies were based on mathematical modeling of fluorescence recovery after photobleaching (FRAP) data, with later studies applying similar approaches to fluorescence correlation spectroscopy (FCS) data, and with the most recent analyses done by single-molecule tracking (SMT)[1,2]. In their *Nature Methods* paper, Gebhardt *et al.*[3] developed a new strategy to improve the signal-to-background ratio in SMT of fluorescently tagged transcription factors in living mammalian nuclei, and they applied their method to measure residence times and bound fractions of the glucocorticoid receptor (GR).

The authors reported a residence time for GR that is almost an order of magnitude longer than that obtained by a FRAP measurement[4], and they concluded that FRAP is inaccurate owing to the many complications involved in the mathematical modeling of such

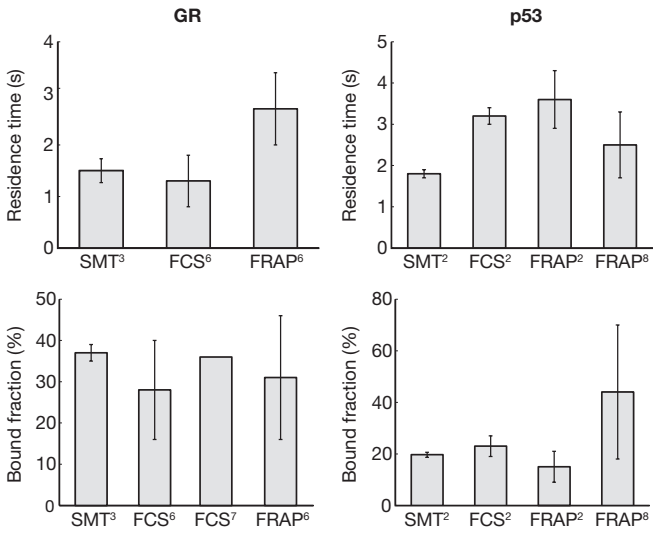


Figure 1 | Consensus in live-cell binding measurements. Consistent estimates for transcription factor residence times and bound fractions on chromatin have now been obtained using three different techniques (SMT, FCS and FRAP) in five different studies (superscripts) applied to two different transcription factors (GR and p53). Bars show published mean values with errors when available (GR, s.d. for FRAP, FCS and SMT; p53, s.e.m. for FRAP and FCS and 95% confidence interval for SMT).

data. Previous studies from other groups have also identified discrepancies in FRAP measurements of chromatin binding that were even larger, exceeding three orders of magnitude[5]. Gebhardt *et al.* cited an early FRAP study of GR[4] that was based on an oversimplification of the photobleaching profile. However, the authors did not cite several later studies, one of which corrected this oversimplification[5] and another that cross-validated the newer FRAP estimate by an FCS analysis of GR binding[6]. These studies yielded residence-time estimates and bound fractions by both FRAP and FCS that are in excellent agreement with those reported by Gebhardt *et al.*[3] by SMT (**Fig. 1**). A second FCS analysis of GR[7] also yielded a bound fraction consistent with the preceding analyses (**Fig. 1**).

We recently reported an analogous consensus for residence times and bound fractions for another transcription factor (p53) in a three-way comparison of FRAP, FCS and SMT[2] (**Fig. 1**). Our p53 estimates were also consistent with an earlier FRAP analysis of p53 (ref. 8). Thus, in stark contrast to earlier measurements[4] and to Gebhardt *et al.*'s conclusions, agreement in live-cell binding estimates has now been obtained for two different transcription factors in five different studies using three complementary approaches. These recent data, acquired using the most current methods for performing FRAP, FCS or SMT, demonstrate that we are reaching consensus on how to extract quantitative binding estimates using not only SMT but also FRAP and FCS.

ACKNOWLEDGMENTS

We thank T. Morisaki, W. Müller, S. Ganguly, C. Zimmer and H. Kimura for comments on the manuscript. D.M. was supported by a Marie Curie international incoming fellowship (GA: 27432). F.M. was supported by the Fondation pour la Recherche Médicale en France (FRM). T.J.S. was supported by the Japan Society for the Promotion of Science (JSPS). J.G.M. was supported by the intramural program of the Center for Cancer Research, National Cancer Institute, US National Institutes of Health.

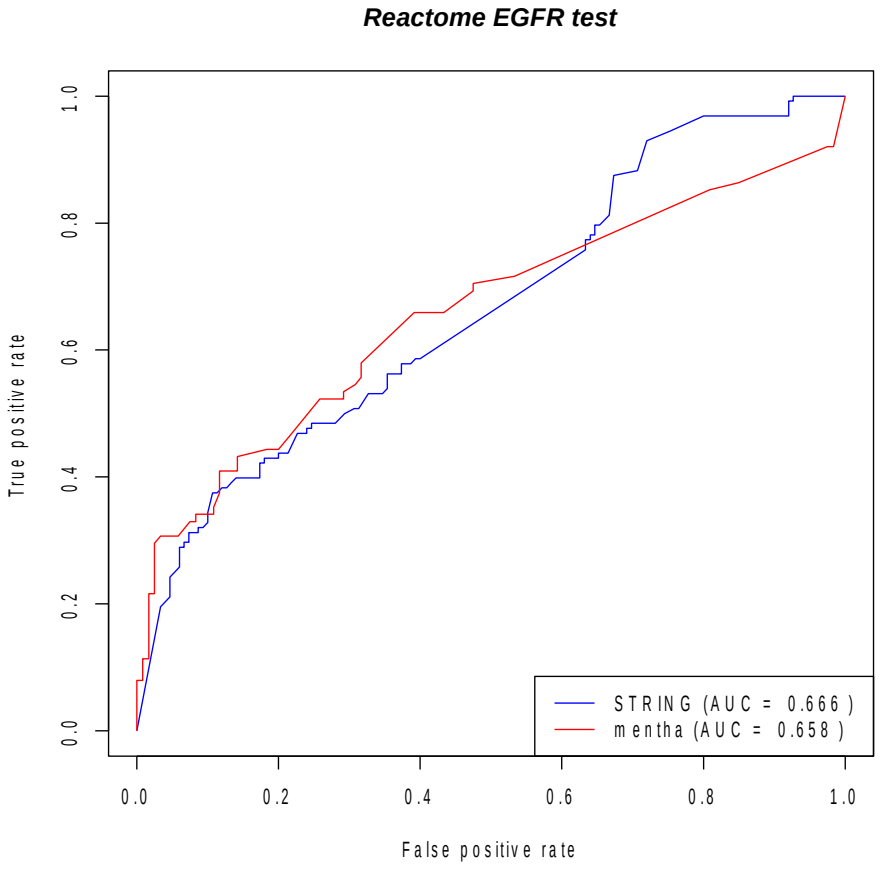
Supplementary Note 1: Considerations About Data Integration

Considerations about curated data integration and inferred data integration

mentha integrates high confidence interaction information curated by iMEX databases. The literature coverage of these databases is not complete. Thus, it is not surprising that some interactions are missing in *mentha*. STRING on the other hand combines interaction information from a larger number of databases, which are curate with different curation rules and curation models, and integrates these data with literature mining and predictions based on a variety of methods. In addition, it uses orthology to transfer interaction information between different organisms. We have carried out an experiment to compare the coverage and precision of the two databases by using as Gold Standard the protein interaction extracted from Reactome, an expert curated pathway database.

To this end, we have downloaded from STRING and *mentha* all the literature-supported interactions occurring among proteins that are annotated as part of the EGF receptor pathway in the Reactome database.

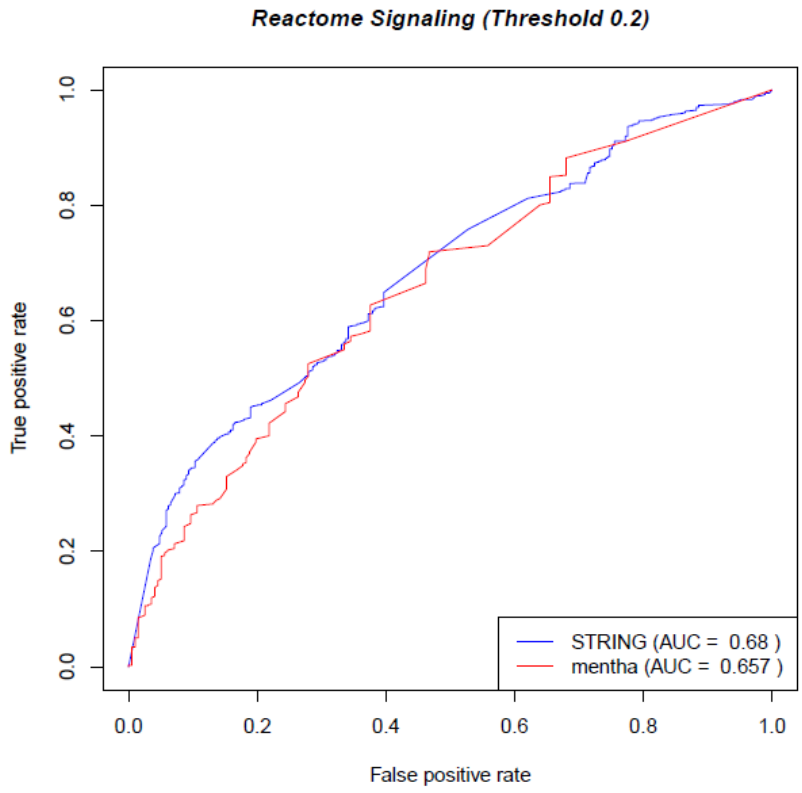
We report here two ROC curves using as positive dataset the list of “direct complex” protein pairs annotated in Reactome and as negative dataset a list of protein pairs that are selected at random between pairs of “EGFR-pathway” proteins. The true positive versus false positive rates of the two ranked lists are plotted.



The analysis reveals that *mentha* does not compare badly with STRING since the two ROC curves have comparable AUC. In addition, as expected, given the difference in integration strategies and scores:

- The *mentha* ROC curve start rising more steeply, suggesting that its score tends to privilege functional direct interaction. In other words, among the interactions scoring high in the ranked list the true/false positive ratio is higher in *mentha* than in STRING consistent with a higher precision.
- On the other hand, the STRING curve recovers towards the end, consistent with a higher coverage.

Despite being instructive, it needs to be pointed out that this type of analysis is unfair with *mentha* because STRING integrates in its database the PPI information curated in Reactome. To further compare the two different integration strategies, we extended this type of analysis to all the "Signaling pathways" in Reactome. The conclusions do not change substantially.



We cannot propose this analysis as a formal comparison of coverage and accuracy in STRING and *mentha*. To perform a proper ROC analysis we should have a trusted "Golden standard". For this purpose, we used the protein interaction dataset compiled in the Reactome website. However, we have no reason to believe that this is any better than the interactome compiled by IMEx database curation or the one assembled by STRING.

Nature Methods: doi:10.1038/nmeth.2561

We went through a manual evaluation of the interactions that are present in the Reactome and STRING dataset but not in *mentha*. Many of the observed discrepancies are a consequence of differences in the curation rules. For instance, a very large fraction of the top ranking interactions in Reactome and STRING are interactions between ubiquitin and ubiquitinated proteins. According to the PSI-MI standard and the IMEx manual, these are considered as post-translational modifications and not protein interactions and are therefore missing from the *mentha* dataset.

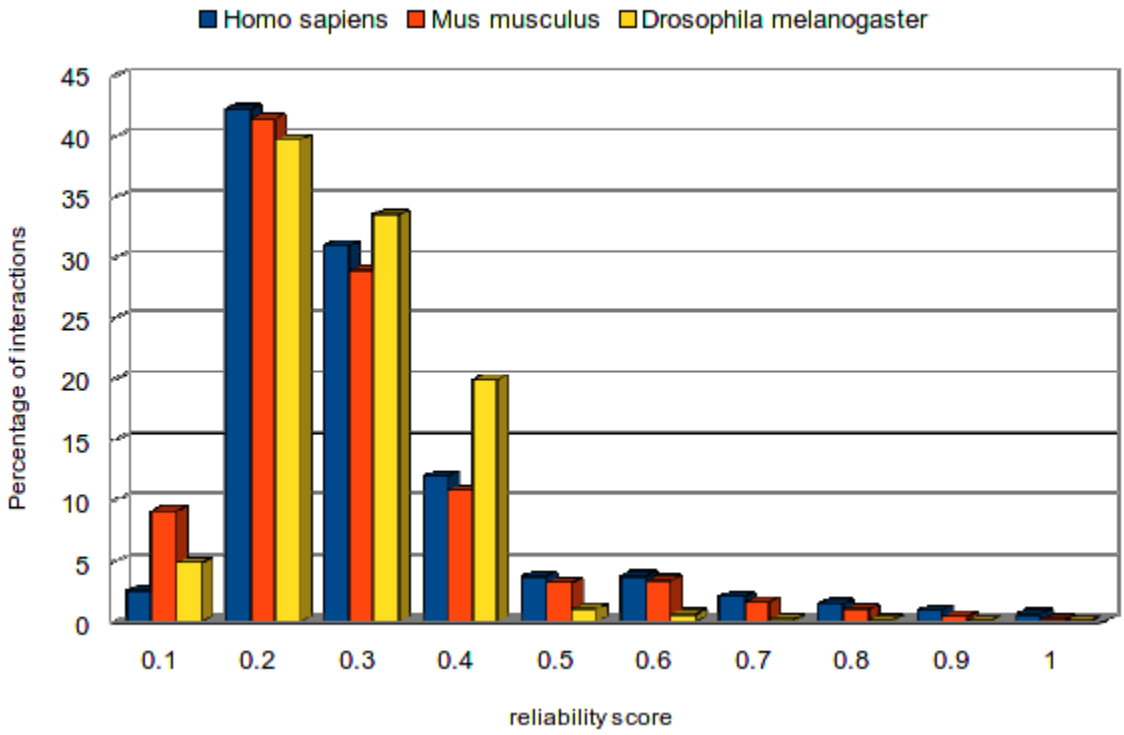
Nature Methods: doi:10.1038/nmeth.2561

Supplementary Note 2: Scoring Function

A score varying between 0 and 1 is assigned to each interaction archived in *mentha*. The score takes into account all the aggregated experimental evidence retrieved from the different databases. The score is calculated, as defined in the MINT database[1], as a function of the cumulative evidence (x) as:

$$S = 1 - a^{-x}$$

a determines the initial slope of the curve. We arbitrarily chose a=1.4 in order to obtain a convenient dynamic range of the score distribution.



The x exponent represents the combined experimental evidence and it is obtained by adding up all the experimental evidence weighted by specific coefficients that consider the type of experimental approach and the size of the experiment:

$$x = \sum_i d_i e_i + n/10$$

Coefficient definition

- **d** reflects the size of the experiment. Experiments are defined large scale if the article reporting them reports more than 50 interactions otherwise they are defined small scale. This coefficient is set to 1 for small scale and to 0.5 for large scale experiments.
- **e** depends on the type of experiment supporting the interaction and emphasizes evidences of direct interaction (e=1) with respect to experimental support that does not provide unequivocal evidence of direct interaction, i.e co-ip, pull down etc (e=0.5).
- **x** takes into account the number of different publications (**n**) supporting the interaction.

Nature Methods: doi:10.1038/nmeth.2561

The MINT scoring function assigns a score close to 1 only to interactions supported by many different reports and experimental approaches while an interaction supported, for instance, by a single high throughput pull down experiment will receive a score of 0.2.

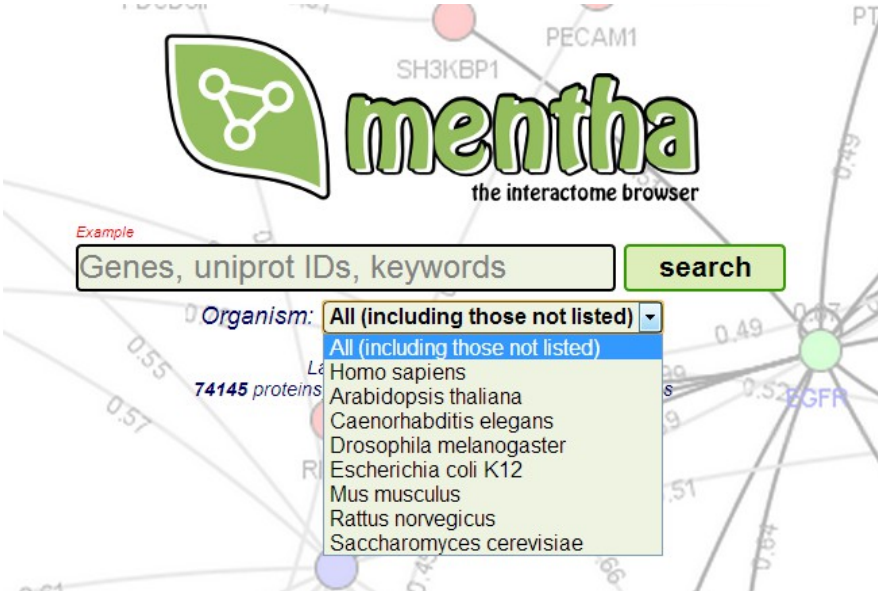
References

1. Licata, L. *et al.* MINT, the molecular interaction database: 2012 update. *Nucleic Acids Res.* **40,** D857–861 (2012).

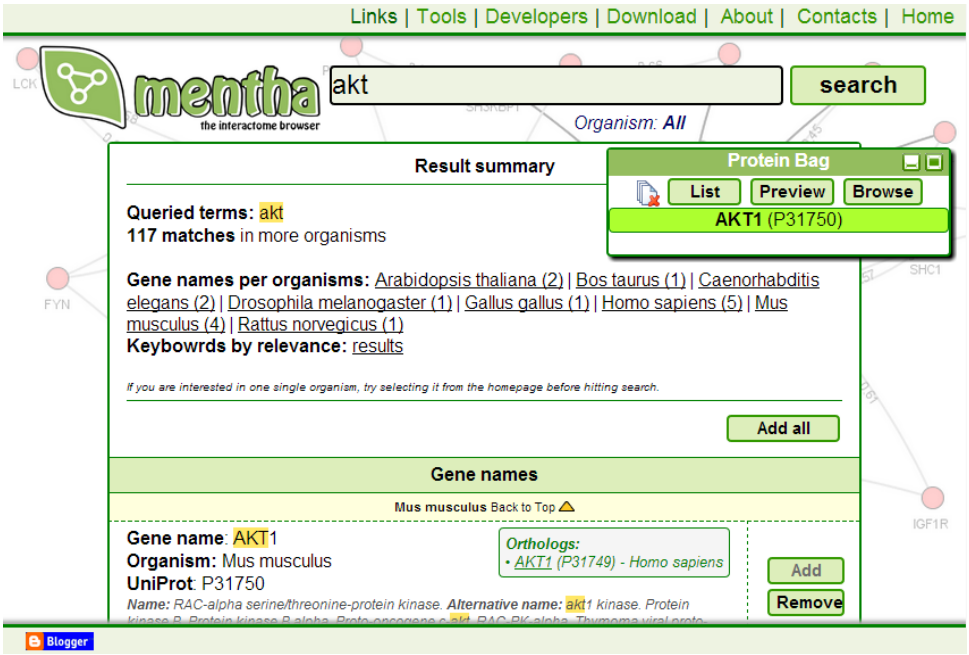
Supplementary Note 3: Quick User Guide

Web site

The integrated PPI data is made accessible through a website that implements a web search engine. From the homepage, it is possible to select the model organism of interest or to browse the entire database. The interface offers the possibility of searching for one or more gene names, UniProt Accession Numbers or keywords such as “kinase”, “membrane” etc.

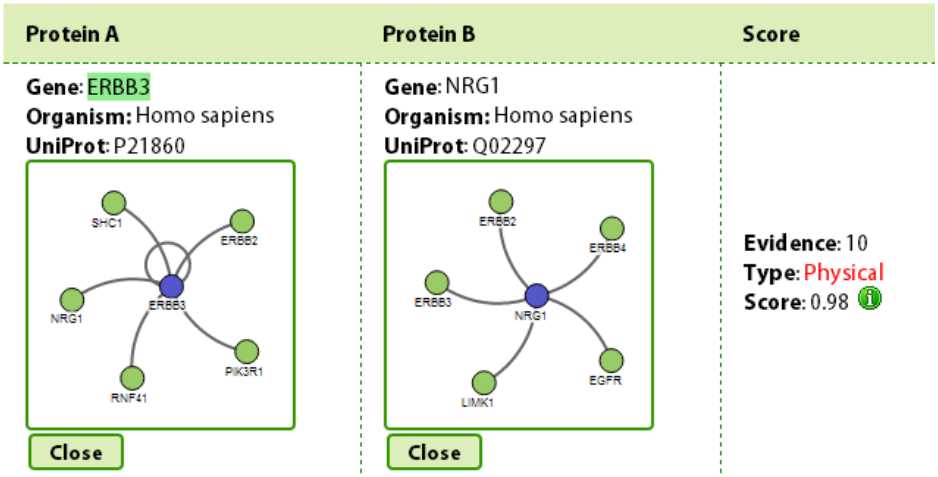


The main advantage of using “All” is that it is possible to find heterogeneous interactions (for example a human protein that was shown to interact with a murine protein). Selecting one model organism allows the user to selectively browse the interactome of the chosen organism as well as offering the possibility of computing paths among proteins in the *interactome browser*. In order to maintain consistence, paths are computable only on specific organisms and not on heterogeneous proteins.



mentha web site offers tools to build a customized protein-protein interaction network. To this end, we have implemented the "protein bag". The user can select one or more proteins from search result-page and add them to the "protein bag". The proteins that have been collected in the bag are retained also after a second search in order to help the user compiling a list of proteins that could be used as a scaffold in the assembly of an interaction network.

By clicking the "List" button the user will be presented with a list containing all the interactions in which the proteins in the "Protein Bag" are involved, associated to their respective scores. For each interaction, the Gene Ontology terms[1] that are common to both partners are listed in order to provide additional evidence for the biological significance of the interaction. By clicking the "show evidence" button all the experiments and papers supporting the interaction are displayed, together with hyperlinks to the relevant literature and to the database that originally annotated the entry. Interactors can be viewed from the List page using "Top 5 Interactors" button.



The interaction page, with associated annotation, can be downloaded as a text file for local use. The "Preview" button allows the user to quickly identify direct interactions, if any, among the proteins in the "protein bag." The extended network, including the selected proteins and their partners, can be visualized via the graphical applet (see below) by pressing the "browse" button.

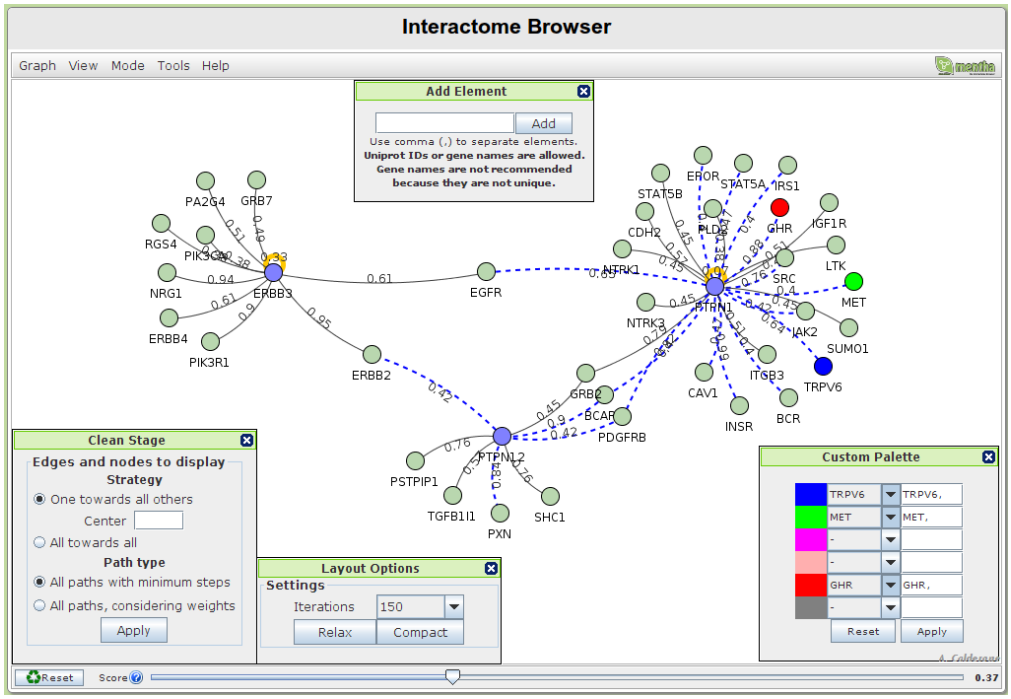
The website also offers advanced graph analysis tools. The first tool "Direct interactions" can retrieve direct interactions between two lists of proteins. A second tool "Subnetwork extractor" allows the user to extract a network starting from a set of proteins that we term network "seed". This tool retrieves partner proteins and partners of partners, allowing the user to connect the "seed" proteins if they are separated by a maximum of three proteins. Finally, "Minimum paths" allows the user to find all the paths connecting the proteins in two protein lists.

The website has been implemented in PHP 5 and JavaScript, with AJAX techniques and jQuery.

Graphical applet

The graphical application developed for this project is intended to help the user to visualize and browse through the entire network, starting from the protein(s) of interest. The graphical application represents a network as a graph and offers a series of tools to mold the network itself. The layout is

obtained using the Fruchterman and Reingold algorithm[2] in order to grant a clean visualization in most circumstances.



The application gives the user the possibility of interactively exploring the interactome. Nodes and edges are sensitive to clicks. Clicking over one edge shows all evidence supporting an interaction. Furthermore, the right-click menu that appears over one node gives the user the ability to operate different functionalities. The first option is to “expand” the network by retrieving interactors of a specific protein. It is also possible to “prune” a node, by removing all the connected nodes that have degree 1. It is possible to delete a node from the stage, together with all its interactors with degree one. The “expand” and “prune” actions can be also applied to the whole network. “Expand and prune” is a special function that does both operations at the same time in order to broaden and clean the network. Finally, the “add element” tool in the “tools” menu allows to add specific proteins as you explore the network. After a series of modifications, it is always possible to reset the network back to its starting configuration. In order to control the layout and the visualization it is possible to color nodes and to fix them so that they can easily be spotted in complex networks. The graphical application also offers a series of tools to compute and show the shortest path, or all the paths, connecting two nodes or one node to a set of proteins; paths are computed by using the Dijkstra’s algorithm[3].

One important characteristic of this application is that it can be easily embedded in webpages by using an <iframe> html tag. It is possible to feed the application with a series of UniProt AC numbers in order to visualize their interactions. This characteristic has already been exploited by the online version of FEBS letters articles, the HuPho[4] database (http://hupho.uniroma2.it) and the DBATE database (http://160.80.34.123/DBATE/).

References

1. Ashburner, M. *et al.* Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* **25,** 25–29 (2000).
2. Fruchterman, T. M. J. & Reingold, E. M. Graph drawing by force-directed placement. *Softw Pr. Exper* **21,** 1129–1164 (1991).
3. Dijkstra, E. W. A note on two problems in connexion with graphs. *Numer. Math.* **1,** 269–271 (1959).
4. Liberti, S. *et al.* HuPho: the human phosphatase portal. *Febs J.* **280,** 379–387 (2013).

Supplementary Methods

The data used for this project is manually curated by the databases adhering to the IMEx consortium[1] (International Molecular Interaction Exchange, http://www.imexconsortium.org). IMEx partners adopt a common curation policy that entails the use of controlled vocabularies (PSI-MI2.5) meant to facilitate data integration. In addition, they implement PSICQUIC[2], a project whose aim is to standardize the access to molecular interaction databases. PSIQUIC grants that data are compatible and that queries are interpreted in the same way by each database. *mentha* is assembled by a merging procedure that runs weekly, and that creates non-redundant data. The merging procedure behind *mentha* uses the standard REST interface.

The data returned by each server is represented as a series of rows. Formally, the data format is called PSI-MI TAB format, where fields are separated by tabulation. Each row contains details about one piece of evidence from one single paper. Each piece of evidence is enriched by a variety of annotations describing the experimental details. The annotations that are retrieved and integrated in the *mentha* resource are: the two proteins involved in the interaction together with their relative taxid (organism identifier), the interaction type, the experimental method used and the publication identification number. Once data have been retrieved, the raw information is processed to identify unique interactions.
We define a piece of evidence as <Protein A; Protein B; Interaction type; Experimental method; Publication>. Any row missing one or more of these elements will be rejected. MINT[3] and IntAct[4] identify most proteins by reporting UniProtKB identifiers, while DIP[5] and BioGRID[6] use different ones. DIP often uses its specific identifiers while BioGRID uses Entrez Gene (GeneID). A GeneID identifies genes and not proteins. GeneID and other non-UniProtAC identifiers are remapped to UniProtAC through a mapping service offered by UniProt available at http://www.uniprot.org/?tab=mapping. Almost all papers are indicated by their PMID but, in case a DOI is used, the procedure uses the DOI to find its relative PMID. The merging procedure analyzes the terms used to describe the interactions (PSI-MI 2.5) by using the API offered by OLS[7,8]. In order to understand whether two entries represent the same piece of evidence curated by different databases, rows that have the same protein identifiers and the same PMID are scanned to analyze whether the differences in interaction types and experimental evidence are the result of different level of curation detail and thus represent the same piece of evidence. Interaction types and experimental methods are used to "climb up" the ontology tree to see if a common parent exists. Only evidence curated at the highest level is retained. This approach is tolerant to different levels of curation and to ontology changes. The "clean data", where identifiers are normalized and genetic interactions and duplicated rows have been removed, are organized in the evidence table and the binary interaction table. Genetic interactions are not considered by this procedure.

We have merged data in the following order: MINT, IntAct DIP, MatrixDB[9] and BioGRID. All rows in the dump are processed and archived in a new table. Fields are split to extract protein identifiers. Evidence without UniProtKB identifiers are normalized through UniProtKB mapping. If an evidence with same partners and same publication is already contained in the evidence table, the procedure checks whether the new row and the already archived evidence are curated at a different level, and thus identical. If the two evidence are classified as identical, the duplicate is removed and

only the entry curated at the highest level of detail is retained. From the evidence table, unique binary interactions are identified and archived separately according to their respective organisms.

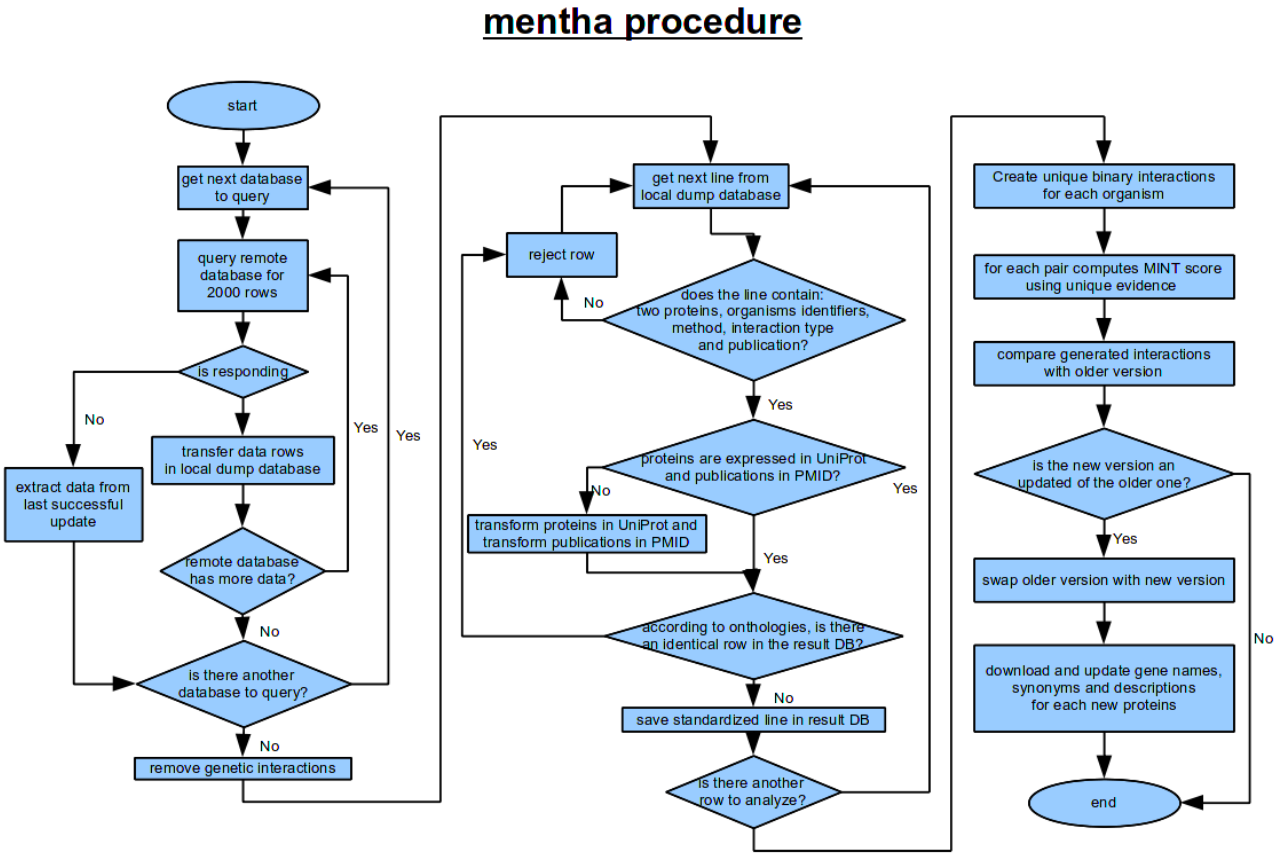


***Illustration 1**: mentha procedure. Flowchart of the key steps of the automatic merging procedure that assembles mentha every week.*

Finally, *mentha* displays orthologies in both the website and the graphical application. Orthologies were calculated using InParanoid 4.1 algorithm from InParanoid7[10] locally, We extracted reviewed proteomes from UniProt for Homo sapiens, Mus musculus and Rattus norvegicus. We adapted InParanoid 4.1 to the latest version of blast-2.2.27 and we used BLOSUM62 as a substitution matrix. We computed orthologies between Homo sapiens against Mus musculus and Rattus norvegicus. Orthologies are archived separately and displayed when the user is using the "All" browsing option, or the OrthoHighlight in the graphical Application.

References

1. Orchard, S. *et al.* Protein interaction data curation: the International Molecular Exchange (IMEx) consortium. *Nat. Methods* **9,** 345–350 (2012).
2. Aranda, B. *et al.* PSICQUIC and PSISCORE: accessing and scoring molecular interactions. *Nat. Methods* **8,** 528–529 (2011).
3. Licata, L. *et al.* MINT, the molecular interaction database: 2012 update. *Nucleic Acids Res.* **40,** D857–861 (2012).

4. Kerrien, S. *et al.* The IntAct molecular interaction database in 2012. *Nucleic Acids Res.* **40,** D841–846 (2012).
5. Salwinski, L. *et al.* The Database of Interacting Proteins: 2004 update. *Nucleic Acids Res.* **32,** D449–451 (2004).
6. Stark, C. *et al.* BioGRID: a general repository for interaction datasets. *Nucleic Acids Res.* **34,** D535–539 (2006).
7. Côté, R. G., Jones, P., Apweiler, R. & Hermjakob, H. The Ontology Lookup Service, a lightweight cross-platform tool for controlled vocabulary queries. *BMC Bioinformatics* **7,** 97 (2006).
8. Côté, R. G., Jones, P., Martens, L., Apweiler, R. & Hermjakob, H. The Ontology Lookup Service: more data and better tools for controlled vocabulary queries. *Nucleic Acids Res.* **36,** W372–376 (2008).
9. Chautard, E., Fatoux-Ardore, M., Ballut, L., Thierry-Mieg, N. & Ricard-Blum, S. MatrixDB, the extracellular matrix interaction database. *Nucleic Acids Res.* **39,** D235–240 (2011).
10. Ostlund, G. *et al.* InParanoid 7: new algorithms and tools for eukaryotic orthology analysis. *Nucleic Acids Res.* **38,** D196–203 (2010).

Nature Methods: doi:10.1038/nmeth.2561