



Sandra Orchard (10/05/12)



May 2012

A tour of the IntAct Portal

<http://www.ebi.ac.uk/intact>

This tutorial will

1. Give you the opportunity to perform a simple search and graphically visualise it
2. Extend the search to produce a more extensive network
3. Look at the details of a individual interaction
4. Perform more complex, directed searches

The IntAct curation policy is to provide the user with all the experimental detail described in the originating paper, with all entries being fully IMEx- [1] and MIMIX-compliant [2] and providing extra levels of detail beyond these minimum requirements. To do so, IntAct makes extensive use of a number of controlled vocabularies, primarily PSI-MI [3] to describe the technical details of the experiment, binding sites, protein tags and mutations and Gene Ontology [4] to describe the subcellular location an interaction may be shown to occur in or the function of an enzyme in an enzyme/substrate assay. Interacting molecules are systematically mapped to stable identifiers from public databases such as UniProtKB for proteins [5], ChEBI for small molecules [6], Ensembl for genes [7] and the DDBJ/EMBL/GenBank nucleotide databases for nucleic acids [8]. Features within a molecule, such as a binding site on a protein, are mapped to the sequence/structure given in the under-lying database and remapped should a new version of the underlying sequence be released. Binding sites are also cross-referenced to the InterPro database [9], whenever possible.



This work is licensed under the Creative Commons Attribution-Share Alike 3.0 License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-sa/3.0/> or send a letter to Creative Commons, 543 Howard Street, 5th Floor, San Francisco, California, 94105, USA.

1. Overview of the portal's layout

This web site uses tabs to give you quick access to the various views of the data

Home: gives access to the Quick and Advanced Search, documentation and data downloads.

Search: allows our users to perform more complex searches.

Interactions: shows the interactions selected by a user query. Binary interactions are shown in a table, in which the user can configure the columns displayed. Various links allow the user to gather more information

Browse: allows users to list molecules interacting in the current set of interaction selected. Users can narrow down the current dataset by browsing the GO ontology and apply a filter by selecting a term. Finally one can link out to other resources using the current set of molecule interactions.

List: allows users to browse the list of interacting molecules by type (protein, small molecule and nucleic acid). A subset of these molecules can be selected and used to select a new set of interactions or link out to other resources.

Interaction Details: shows the full details that were captured by our curators, such as: textual annotation, more cross references, the complete list of participants (as opposed to spoke expanded interaction), binding domains and other experimental features such as sites of mutational analyses.

Graph: shows the current interaction network in our simple viewer and gives the option to users to open their query in Cytoscape, thus allowing more interactive manipulation of interaction networks.

2. A simple search and visualisation of data

a. Using the Quick Search

In this search panel you are free to type anything that might relate to interactions, whether it is properties of their interactor (gene name, Accession Numbers, GO term...) or more specific to the interaction such as publication ID, authors, experimental detection method, ... Usually users will type an accession number or gene name.

In this exercise you will **perform a very simple search and look at your results in the IntAct viewer**. You may try **ONE of the following Accession numbers listed below**, or alternatively, you may try a protein you are interested in through your own work. However, you may find there is little or no data for your protein in the database, particularly if it originates from a non-model organism which are much less well studied.

P01375 (TNFA_HUMAN)

At5g03455 (Arabidopsis locus name, CDC25)

yjhG (E.coli gene name)

The initial search yields information on potential binary interactions made by your protein.

b. Extending your search

Your current network shows you some proteins of interest all of which interact with the molecule for which you made the initial search. You may wish to extend the Search and extend your network.

Open the 'Lists' tab. In the default view, you will see listed all the proteins in your current search (nucleic acids and small molecules are listed in the other tabs in your view).

Select any 3 of the molecules in the list and use the 'Search Interactions' button to find all the interactors with these molecules and add them to your interaction network (again, be warned – not every molecule you search with will have additional data associated with it in the database).

c. Visualising your search

Visualize the currently selected set of interactions by opening the Graph tab. This will give you a simple, immediate representation of your interaction network. Currently, this functionality is only available if you have up to 300 interactions but we are working on extending (possibly lifting altogether) this limitation.

Taking it further using the Cytoscape integration tool

You will find on the right hand side panel the Cytoscape icon that, once clicked, will load Cytoscape using Java Web Start technology (which should be enabled on your computer to work properly).

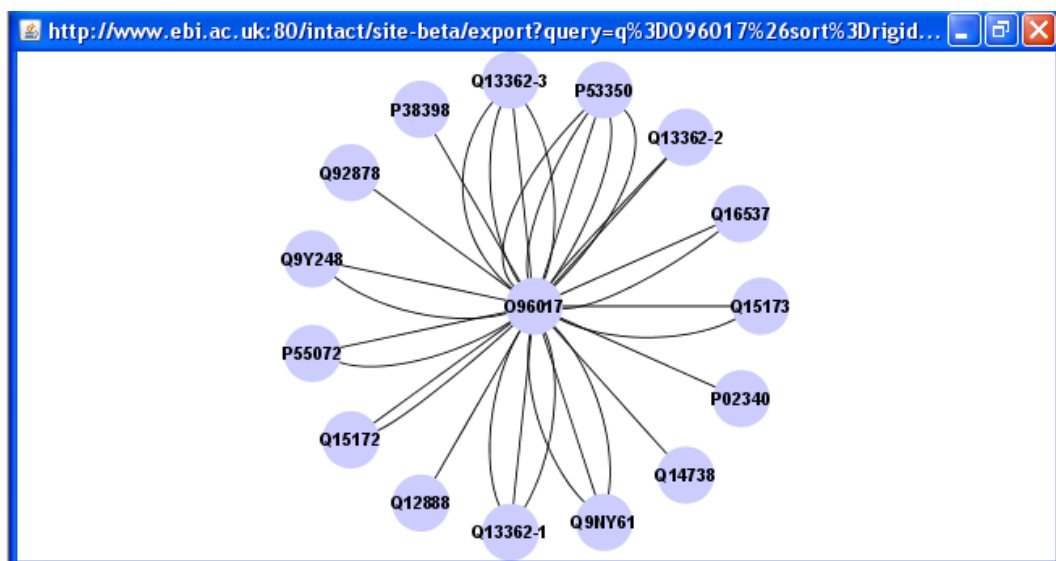
Once Cytoscape has loaded, your dataset is then imported and your network displayed. Please refer to Cytoscape documentation for more information about using this powerful tool.

Open your current interaction network in Cytoscape.

On the left hand side of the network are 4 tabs (Network, VizMapper, Editor and Filters), select the VizMapper, under the drop down list 'Current Visual Style' choose 'Sample 1'.

Hint: You can also beautify the network by applying a layout to it.

Example: Layout > yFiles > Circular



To add extra information to your network, the interaction edge can be given a different colour depending on the *detection method used to identify the interaction*.

You can achieve this effect by following these steps:

Step 1. In the Visual Mapping Browser (left box), expand the Edge Color node.

Step 2. Select *detection method* as the value for the Edge Color node.

Step 3. Select *discrete mapping* as the value for the Mapping type. You should see the list of detection methods below.

Step 4. You could choose your favourite colours for the detection methods, but Cytoscape has an easy way to assign different colours to the values. To do so, right-click on top of discrete mapping and click on Generate Discrete Values > Rainbow 1. You can use this powerful system to show labels, change colours and generate a beautiful network. It is also possible to merge interaction networks or overlay other data, such as micro-array data with the Cytoscape tool.

3. Exploring a single interaction in more depth

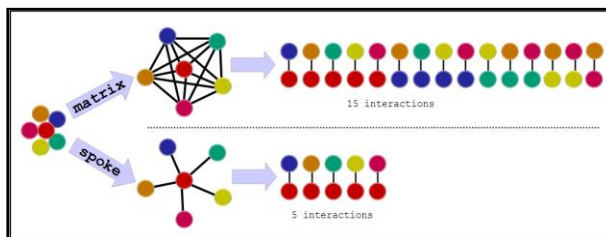
Return to the IntAct home page. Perform a Quick Search using UniProtKB accession number **P06239**

In the list, find the interaction between lymphocyte specific protein kinase (lck) and Mediator complex subunit 28 (Med28)

Q1 How many methods has this interaction been detected by?

Q2 How many publications has this data been taken from?

You will see a list of binary interactions in the table shown, irrespective of whether the data was taken from a binary Y2H or an n-ary affinity chromatography experiment. Whenever the data was reported as a co-complex involving more than two molecules, we store it as such in the IntAct database and post-process it so the portal can show it as binary interaction. This post-processing is the Spoke Expansion model (i.e. the bait is separately paired to each prey):



Q3. Looking at the detailed view for interaction EBI-1206908 – what other tyrosine-specific protein kinase can be found in complex with Lck-Med28?

Return to the ‘Interaction Tab

Downloading the data into Standard formats

In the header of the interaction table you will find a drop down list that contains all the formats currently supported when downloading the interaction data. Select one of them and click the export button next to the list. Please note that PSI-MI XML is only available when the interaction set is no bigger than 1000 interactions.

Refining your search

a. Using the Advanced Search

Clicking on the “Fields” button to the right of the Quick Search box will open up the Advanced Search, allowing you to specify one or more fields you wish to search in, and building the query for you as you progress.

For example, you may wish to collect together all the

Click on “Fields” and select “*Organism*” from the Pulldown menu – type in *Human* as your organism.

Further refine the search by adding “*Interaction detection method*” as

“*Experimental*” – you should see a slight drop in interaction number on the Interaction tab as some inferred data is filtered out. Finally, actively filter out all the two hybrid data by checking the NOT box and selecting “Interaction *Detection method*” as “*two hybrid*” – you should lose about 30,000 interactions as all the two hybrid (and child thereof) data is removed.

If you want to construct more complex queries we recommend you take a look at the Molecular Interaction Query Language, accessible from the quick search panel. This will allow you to write more complex queries, for example:

Repeat the query *species:human* This will identify all the interactions where one of the participants has ‘human’ in the name. Take a note of the number of hits.

Now try the query: *species:9606* This will select all interactions where at least one of the participants is of human (NCBI Taxonomy ID = 9606) origin. Why is there a drop in the number of interactions observed. ‘Clear’ the current search.

Now try the query *taxidA:9606 AND taxidB:9606*

This query selects all interaction where BOTH interactors are of human origin. You will see the numbers differ significantly. This is due to mixed species interactions. Some of these are biologically significant (e.g. human – HIV proteins) but others are caused by the experimental systems constructed by bench scientists e.g. a Flag-tagged human protein expressed in a mouse cell line and used to co-immunoprecipitate mouse proteins.

b. Using the Ontology Search

Open the Search Tab. This panel is specialised to give you an easy access to ontology search. So far you can search on 4 ontologies:

- Gene Ontology
- InterPro
- PSI-MI
- ChEBI

Whenever you start typing a query in this search panel, the system will search as you type and propose a list of matching controlled vocabulary terms. You can then select one of them and select matching interactions.

Type: *mitosis* in the Ontology Search box.

You will be presented with a few choices, please note that each term is followed by the count of matching interactions in the IntAct database.

Select the parent term ‘mitosis’ (GO:0007067) using the keyboard cursor keys, complete the search and you will be taken to the interaction tab. This now gives you ALL the interactions for proteins in IntAct which GO have annotated as being involved in the process of mitosis. Add the term ‘*AND species:human*’ to limit this to interactions in which one of the interactors is of human origin.

Return to the Search Tab

c. Extending your search via PSICQUIC and IMEx

At the same time that you perform a query in IntAct, we simultaneously also perform the same query across all the databases hosting a PSICQUIC server and also across the consistently curated, non-redundant set of interactions available via the IMEx website.

Go to the IntAct home page and perform a quick search on ‘*brca2*’. You will see a response similar to that listed below.

- > 93 binary interactions were found in IntAct. 23 of them are originated from [spoke expanded co-complexes](#) and you may want to [filter](#) them.
- > Your query also matches 854 interaction evidences from 7 other databases.
- > Your query also matches 23 interaction evidences from 1 other IMEx databases.

Clicking on the hyperlink will allow you to access additional data. Remember, collectively the PSICQUIC databases contain highly redundant data and also predicted and inferred data as well as experimentally curated. The IMEx set will give you additional data curated to the same high standards as is IntAct, and this dataset is non-redundant with regard to experimental evidence.

cd. Limiting the scope of search result using the filter panel

Now that you know how to search for molecular interaction data we will see how to limit the scope of your searches. After the preliminary search, your results are a mixture of experimental binary data and binary data derived by performing a spoke expansion of co-complex data. IntAct allows you to filter the expanded binaries out of your final dataset, should you wish.

Ticking the corresponding boxes, you will on search within these categories. Return to the IntAct Home Page.

Try the query: *species:yeast AND detmethod:"Affinity purification"*

Under the heading tabs, there is a statement

- > n binary interactions were found. n of them are originated from [spoke expanded co-complexes](#) and you may want to [filter](#) them.

Click on [filter](#) and see the effect on your final number of interactions.

4. Putting your data into context

IntAct allows you to access to more content, based on the currently selected set of interactions. Please note that some of these functionalities will only allow you to include up to 200 molecules - if you exceed this number you will see the warning icon ⚠. Note – you will need to have the Pop-up window enabled for this to work.

Search on the UniProt identifier *CHK2_HUMAN*

a. Listing the molecules involved by specific type

Clicking on the ‘List All’ present under each molecule type will open the corresponding list of molecule in the List Tab.

Open the Protein tab and select ALL interactors and visualise them using **Reactome SkyPainter**. This will show you the pathways in which these molecules are known to play a role. Which pathways are most heavily represented in this interaction network?

Return to the Browse tab and use the Gene Ontology to see if the Biological Process that these proteins are annotated to reflects the results you obtained from Reactome?

You may also visualise

Proteins by InterPro domain: Opens the InterPro domain search and shows in a single display the proteins interacting in your interaction set.

Proteins by Chromosomal location: Sends your list of proteins to Ensembl’s Karyotype viewer and overlays the proteins on the chromosomes.

Proteins by mRNA expression: Sends your set of proteins to the ArrayExpress Atlas that will show the known gene expression based on experimental studies.

References

References

1. Orchard, S., Kerrien, S., Jones, P., Ceol, A., Chatr-Aryamontri, A., Salwinski, L., Nerothin, J., Hermjakob, H. (2007) Submit your interaction data the IMEx way: a step by step guide to trouble-free deposition. **7 Suppl 1**, 28-34
2. Orchard, S., Salwinski, L., Kerrien, S., Montecchi-Palazzi, L., Oesterheld, M., Stümpflen, V., Ceol, A., Chatr-aryamontri, A., Armstrong, J., Woollard, P., et al. (2007) The Minimum Information required for reporting a Molecular Interaction Experiment (MIMIx) *Nat. Biotechnol.*, **25**, 894-898
3. Kerrien, S., Orchard, S., Montecchi-Palazzi, L., Aranda, B., Quinn, A.F., Vinod, N., Bader, G.D., Xenarios, I., Wojcik, J., Sherman, D., et al (2007) Broadening the horizon-level 2.5 of the HUPO-PSI format for molecular interactions. *BMC biology*, **5**, 44
4. Blake, J.A., Harris, M.A. (2008) The Gene Ontology (GO) project: structured vocabularies for molecular biology and their application to genome and expression analysis. *Current protocols in bioinformatics*, **7**, 7.2
5. The UniProt Consortium (2011) Ongoing and future developments at the Universal Protein Resource.. *Nucleic acids research*, (39), d214-219
6. de Matos P , Alcántara R , Dekker A , Ennis M , Hastings J , Haug K , Spiteri I , Turner S , Steinbeck C (2011) Chemical Entities of Biological Interest: an update. *Nucleic Acids Res* 38:d249-254
7. Flicek P , Aken BL , Ballester B , Beal K , Bragin E , Brent S , Chen Y , Clapham P , Coates G , Fairley S , Fitzgerald S , Fernandez-Banet J , Gordon L , Gräf S , Haider S et al (2010) Ensembl 2009. *Nucleic acids research* (38):d557-562
8. Tateno, Y. (2008) International collaboration among DDBJ, EMBL Bank and GenBank. *Tanpakushitsu kakusan koso. Protein, nucleic acid, enzyme* **53**, 182-189
9. McDowall J , Hunter S (2011) InterPro protein classification. *Methods Mol Biol* 694:37-47

Further Reading

Aranda B , Achuthan P , Alam-Faruque Y , Armean I , Bridge A , Derow C , Feuermann M , Ghanbarian AT , Kerrien S , Khadake J , Kerssemakers J , Leroy C , Menden M , Michaut M , Montecchi-Palazzi L , Neuhauser, S.N. Orchard, S., Perreau, V., Roechert, B., van Eijk, K. and Hermjakob, H. (2010) **The IntAct molecular interaction database in 2010**. *Nucleic acids research* 38, d525-531

Aranda B , Blankenburg H , Kerrien S , Brinkman FS , Ceol A , Chautard E , Dana JM , De Las Rivas J , Dumousseau M , Galeota E , Gaulton A , Goll J , Hancock RE , Isserlin R , Jimenez RC , Kerssemakers J , Khadake J , Lynn DJ , Michaut M , O'Kelly G , Ono K , Orchard S , Prieto C , Razick S , Rigina O , Salwinski L , Simonovic M , Velankar S , Winter A , Wu G , Bader GD , Cesareni G , Donaldson IM , Eisenberg D , Kleywegt GJ , Overington J , Ricard-Blum S , Tyers M , Albrecht M , Hermjakob H. (2011) PSICQUIC and PSIScore: accessing and scoring molecular interactions. *Nat. Methods* 8: 528-529

Orchard, S., Kerrien, S., Abbani, S., Aranda, B. Bhate J , Bidwell S , Bridge A , Briganti L , Brinkman, F.S. , Cesareni G., Chatr-aryamontri A , Chautard E , Chen C , Dumousseau M , Goll J Hancock, R.E.W., Hannick, L.I., Jurisica, I., Khadake, J., Lynn, D.J., Mahadevan, U., Perfetto, L., Raghunath, A., Ricard-Blum, S., Roechert, B., Salwinski, L., Stümpflen, V., Tyers, M., Uetz, P., Xenarios, I., Hermjakob, H. (2012) Protein interaction data curation: the International Molecular Exchange (IMEx) consortium. *Nat Methods*, 9, 345-350