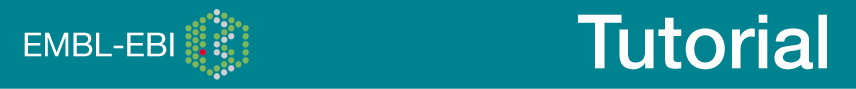
**INSILICO SYSTEMS BIOLOGY PRACTICALS**

**EXPERIMENT-21**



Sandra Orchard/Pablo Porras (V6 01/07/13)

A tour of the IntAct Portal

[*http://www.ebi.ac.uk/intact*](http://www.ebi.ac.uk/intact)

*With this tutorial you will learn how to:*

* *Perform a simple search in IntAct and graphically visualise it*
* *Extend the search to produce a more extensive network*
* *Look at the details of an individual interaction*
* *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- and MIMIx-compliant 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 to describe the technical details of the experiment, binding sites, protein tags and mutations and Gene Ontology 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, ChEBI for small molecules, Ensembl for genes and the DDBJ/EMBL/GenBank nucleotide databases for nucleic acids. 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, whenever possible.*

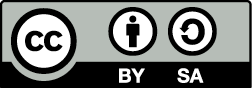
A simple search and visualisation of data

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 perform the exercise 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.

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

 Type ‘CHK2’ into the IntAct Quick search and hit ‘Search’

**Question 1** How many binary interactions does this find – and how many is this reduced by when you ‘filter’ out the spoke expanded co-complex data?

 Graphically view the data within IntAct

 Open the ‘List’ tab and select all proteins (use the check box at the top of the table) and use ‘search for interactions’ to perform a second round of search.

 **Question 2** How many interactors can you now find?

 Clear the search and perform a second search on CHK2\_HUMAN

 **Question 3** How many interactors do you now find?

 Go to the detailed view of the interaction of CHK2 with BRCA1 (EBI-1263190)

 **Question 4**: Which other proteins are co-complexed with this molecule?

 **Question 5**: What protein feature was observed for CHK2 (Open red F) – and on which amino acid is it seen?

# Refining your search

1. **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.

Let’s do a blank search defined only by the field you want to search with. In this example, we will start by looking for all the interaction involving human proteins that can be found in IntAct.

 Clear your search before you start this part of the tutorial.

Click on ‘Fields’ and select ‘Organism’ from the pull-down menu – type in ‘Human’ as your organism.

 Further refine the search by adding “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 “Detection method” as “two hybrid” – you should loose about 10,000 interactions as all the two hybrid (and child thereof) data is removed.

2

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:

Clear your search before you start this part of the tutorial.

Try the query: taxidA:9606

This will select all interactions where at least one of the participants is of human (NCBI Taxonomy ID = 9606) origin. Make a note of the number. ‘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 a 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).

# 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.

 Clear your search before you start this part of the tutorial.  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 mitotis. Add the term ‘AND species:human’ to limit this to interactions in which one of the interactors is of human origin.

3

# 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.

 Clear your search before you start this part of the tutorial.

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



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.

4

Further Reading

## The UniProt Consortium “Reorganizing the protein space at the Universal Protein Resource (UniProt).” Nucleic Acids Res. (2012) 40:71-75

* Leinonen R, Nardone F, Zhu W, Apweiler R “UniSave: the UniProtKB sequence/annotation version database.” Bioinformatics (2006) 22:1284- 1285

## Suzek BE, Huang H, McGarvey P, Mazumder R, Wu CH “UniRef: comprehensive and non-redundant UniProt reference clusters.” Bioinformatics (2007) 23:1282-8

* Mulder NJ et al “New developments in the InterPro database.” Nucleic Acids Res. (2007) 35:D224-8

## Kerrien S et al “The IntAct molecular interaction database in 2012.” Nucleic Acids Res. (2012)40:D841-846

5

IntAct

Answer 1: 43 – reducing to 35

Answer 2: 514 (494 proteins, 6 compounds, 5 nucleic acids and 9 genes) Answer 3: 40

Answer 4: RAD50, TP53BP1 and GINS2

Answer 5: Phospho-threonine – residue 68

6