Protein Networks and Systems Biology
ELIXIR 2015, Bologna, Italy

Biological Networks

Alberto Calderone
sinnefa@gmail.com

Bioinformatics and Computational Biology Unit
Molecular Genetics Laboratory
University of Rome Tor Vergata, Italy

December 11, 2015

Introduction

This training session will be an exploratory session. We will explore different online resources that collect the different data types introduced in this course. We will compare protein interaction, signalling interaction and reaction information.

Protein Interaction Resources

Introduction

In this training session we will go through the steps needed to collect information about protein-protein interactions. We will see how to explore and extract such data, the different resources available and the importance of data integration.

Step 1: Protein Interaction Resources

Browse and compare each of the following resource. Search for your favourite proteins or genes (Examples are: EGFR, EGF, ALB1, Diablo, try also full names like "Epidermal growth factor receptor").

Resource	Web **site**
MINT	http://mint.bio.uniroma2.it/mint/Welcome.do
IntAct	http://www.ebi.ac.uk/intact/
DIP	http://dip.doe-mbi.ucla.edu/dip/Main.cgi
MatrixDB	http://matrixdb.ibcp.fr/
BioGRID	http://thebiogrid.org/

Ask yourself the following questions:

- Could the resource find your protein(s) or gene(s)?
- Could you find the same identical information in more than one resource? (Same papers, same experiments, same annotations...)
- Could you download the results obtained to process them off-line?

This step is to show you some of the online data sources and that you can use them to collect data relevant to your research. You should have noticed that each resource has its own way of offering data to the user and that there are pros and cons to each of them. In general, it is advisable to search more resources.

At this point it should come spontaneous to wonder how we can collect as much information as possible in order to build a network to study. In the following steps we will see what has been done to allow researchers to download as much data as possible in the easiest way, no matter the resource used. In a following training session we will see how to analyse networks in practise.

Step 2: A Common Protocol

Despite having different web interfaces, all the protein interaction databases listed is Step 1 share the same data protocol. This means that they allow external queries to be structured in the same way and that they return results in the same format. This standardisation allowed the development of a unified interface which is available at:

http://www.ebi.ac.uk/Tools/webservices/psicquic/view/

Search again the same proteins your searched in Step 1 and explore the different results and resources listed. Pay attention to the resources you decide to explore because not all the databases listed contain protein-protein interaction information, nor they use the same identical way for representing some data (e.g.: Protein complexes expansion models). Narrow your browsing to the resources listed in Step 1 since they are fully compliant one another.

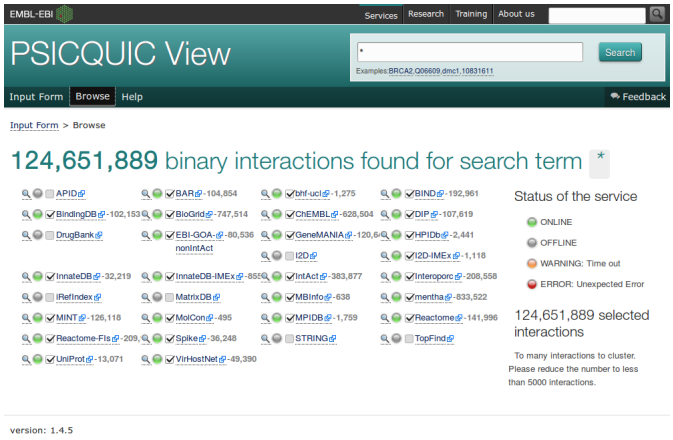

Figure 1: *PSICQUIC View. PSICQUIC View simultaneously searches in various resources. Pay attention to the resources listed because not all resources contain protein-protein interactions and not all resources have the same data representation policies.*

PSICQUIC View web interface translates your query into a common format that is accepted by each resource. Also, the results from each database are structured in the same way. Queries and results follow the PSICQUIC protocol.

For those interested in the protocol details and implementation you can browse to this web-site:

https://github.com/micommunity/psicquic

List of resources supporting PSICQUIC protocol

If you want more details about the various web resources that support the PSICQUIC protocol you can go to:

http://www.ebi.ac.uk/Tools/webservices/psicquic/registry/registry?action=STATUS

You can compare "tags" attributed to each resources and see their differences.

Step 3: Integration Leads to Knowledge

We saw that it is possible to exploit the PSICQUIC View to simultaneously query various databases. Despite this advantage, data is still scattered in different resources and need to be analysed line by line.
We will now see how these data can be merged to ease information accessibility and to score interaction reliability. *mentha* is a web resource that extracts data from different resources to improve coverage and interaction robustness. Differently from PSICQUIC View *mentha* merges results by default facilitating evidence exploration. It also computes an interaction scores based on the number and kind of experimental evidence. Browse to:

http://mentha.uniroma2.it/

When querying *mentha* you can restrict your reach to a specific organism. In the following image we are searching EGFR in the human interactome only. Figure 2.

Search again the same proteins you searched in Step 1 and 2. Try to collect them in the *Protein Bag* and to visualise the interaction network to see if and how your proteins are connected. If you have a correctly configured Java plugin in your browser you can also start the Advanced Network Explorer and perform advanced actions like, path discovery, colouring, network expansion etc... While collecting your proteins of interest, explore the *Results Page* to see how pieces of evidence are grouped and interaction scores (Figure 3).
mentha is an example of how you can collect more data to create a large network of interactions. You can use *mentha* through PSICQUIC or you can download already integrated data in different ways, either via the dedicated download page or from your results page.

Figure 2: *mentha main page. The drop down menu allows the user to restrict the query to main model organisms. Other organisms are available selecting "All".*

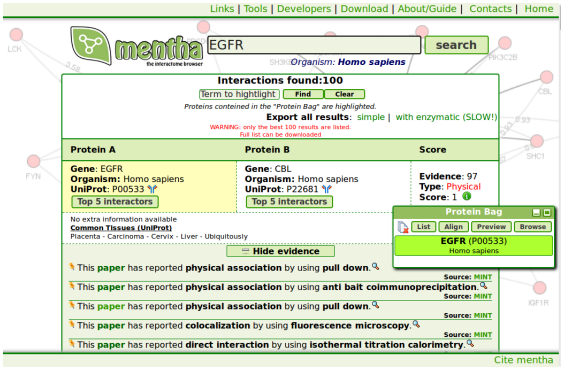

Figure 3: *Results Page. Here you can see interactions listed as pairs. Each interaction can either be "Enzymatic" or "Physical" and it has a reliability score assigned. Below each pair, it is possible to expand the list of evidence that support the interaction. Each piece of evidence is linked the the original paper and, on the right, to the source database, e.g. in this figure MINT.*

Causal Interaction Resources

Introduction

In this section we will get and idea about causal interaction information (or Signalling). This kind of interactions can be seen as "Causal Interactions" since we know what the effect on an interactions is. Differently from Protein-Protein interaction information, causal information is still not well standardised and web resources are still limited. We will see how to explore such data from the largest resource currently available SIGNOR.

SIGNOR: *SIGnaling Network Open Resource*

The core of SIGNOR is a collection of manually-annotated causal relationships between human proteins participating in signal transduction. Other entities annotated in SIGNOR are complexes, chemicals, phenotypes and stimuli. Each entry is associated to the post-translational modifications that cause the activation/inactivation of the target proteins. Also, modified residues causing a change in protein concentration or activity have been curated and linked to the modifying enzymes. Additional modifications such as ubiquitynation, sumoylations, acetylations and their effect on the modified target proteins are also annotated. Navigate to this website:

http://signor.uniroma2.it/

Search again the same proteins you searched in "Protein-Protein Interaction Information". Differently from the previous activity, you can now also search for chemicals and molecules such as "AMP" or "Metformin".
Like for protein interaction information you would be interested in downloading data to perform computational analysis. SIGNOR allows you to export interactions as pair or as SBML model, which are two formats that we will explore in the next sessions of this course.

PPI and Signaling Interactions

Now that you know how to use *mentha* and SIGNOR try to search for the same proteins in both resources and compare and contrast the information you can get from Result Pages of the two resources.

Protein-Protein Interaction and Causal Interaction information

The community working on curating and archiving protein interaction information is active and policies and protocols adopted are mature enough to offer a good service no matter the resource you decide to use; notably the IMEx consortium has been working on this direction for many years. Not only these data is easily accessible by lay users but, thanks to its maturity, also programmatic access is close to be standardised across resources.

On the other hand, causal interaction (or signalling interaction) information is still sparse and policies are not yet defined clearly to allow common access and data exchange. For the sake of completeness we report that there are at the moment two major databases that attempt to offer such data as large network of interactions. One is SignaLink and the other one is SIGNOR, which you got to know in the previous section.

Reaction-Based Online Resources

Lastly, we will explore online resources that archive reaction networks. This kind of interactions are more linked to biochemical effect than to physical interactions. Navigate to this website:

http://www.reactome.org/

On the right side of Reactome's homepage you can find a small search bar.

- Type in the protein name "EGFR". This protein is a receptor involved in signal transduction.
- Since EGFR has multiple matches reactome will ask you to select one alternative, select *Homo Sapiens*.
- Under *Locations in the PathwayBrowser* select *Signal Transduction(Homo sapiens)*
- Among the various sub networks shown, double click the top-left most box *Signaling by EGFR*
- Zoom-in the reactions shown and locate the reaction in which EGFR and EGF are involved. You can find this reaction at the top-left corner of the figure. Click on this interaction.

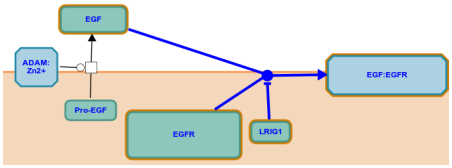

Figure 4: *This figure shows the reaction in which EGFR is involved.*

- Discuss on what are the differences between this reaction and the interactions you found in *mentha* and SIGNOR. What are these entities and why there are non-binary interactions.
- Explore the details given in the lower part of the web page. What are the compartments in which we can find this interactions?
- Try navigating the network using *Following Events*.

Navigate to this website:

https://www.ebi.ac.uk/biomodels-main/

if you search again EGFR and click on the first result, can you guess what are the differences between the data you can retrieve from Reactome and from BioModels?

Conclusions

At the end of this training session you should be aware of different data sources depending on the kind of data you need to process. You should now know where to find protein interactions, signalling interactions and biochemical reactions and you should be aware of the different information that these three different data type convey. You are now also aware of different data curation strategies and the importance of data integration. Data retrievable from these resources is the input data of the following training sessions. We will learn how to analyse protein interaction network, signalling networks and dynamical models.