Tutorial: UniProtKB

GW School of Medicine & Health Sciences (January 28, 2020)

i) What is UniProtKB?

UniProtKB (https://www.uniprot.org/) The UniProt Knowledgebase, the centerpiece of the UniProt Consortium's activities, is an expertly and richly curated protein database, consisting of two sections called UniProtKB/Swiss-Prot and UniProtKB/TrEMBL. It is the central hub for the collection of functional information on proteins, with accurate, consistent, and rich annotation.

ii) Using UniProtKB to answer a scientific question.

Background: Severe acute respiratory syndrome (SARS) is a viral respiratory disease of zoonotic origin caused by the SARS coronavirus (SARS-CoV). Spike (S) proteins of coronaviruses, including the coronavirus that causes severe acute respiratory syndrome (SARS), associate with cellular receptors to mediate infection of their target cells.

In this tutorial, a Simple Protein Search will be performed to answer the following question:
Which proteins of the viral host(human) can interact with and bind to spike glycoprotein of human SARS coronavirus?

Step-by-step instructions:

1. Visit the uniprot.org using browse.

2. Once on the homepage, access the UniProtKB from the top navigation bar on the landing page, -> click on UniProtKB, or directly click on UniProtKB on the interface. (Figure 1)

Figure 1 Access the UniProtKB

3. In order to query for viral host protein, which can interact with and bind to spike glycoprotein of human SARS coronavirus. For this
   a. Click the second tab Advanced from the top navigation bar. (Figure 1)
   b. From the Advanced search dialog box, select Protein name from the drop-down list, type in: spike glycoprotein in the Term search bar. -> On the second line, select AND from the drop-down list, and select Reviewed from the drop-down list. -> On the third line, select AND from the drop-down list. and select Organism from the drop-down list, type in: Human SARS coronavirus in the Term search bar. -> click on Search. ->A list with one protein found and Entry is *P59594* shown below ((Figure 2)

Figure 2 Advance search for spike glycoprotein of human SARS coronavirus

c. For the one protein that was found (SPIKE_CVHSA), ->click on its UniProtKB Entry *P59594*-> navigate to detailed information for the selected protein. Under an Entry menu, there is a list of sub-menus(which can be check or uncheck) ->click Interaction.

d. Scroll to Protein-protein interaction databases, ->Click the hyperlink *P59594*, ->Navigate to a details page(Figure 4), which includes binary interactions with spike glycoprotein. ->
There is one human protein reported, and its description is Angiotensin-converting enzyme 2.-> Click on UniProt hyperlink->Navigate to details page in UniProtKB, which includes information on proteins(Angiotensin-converting enzyme 2). (Figure 5) -->Click on Graph the top navigation bar on the IntAct's home page, There is Network visualization of two proteins. (Figure 6).

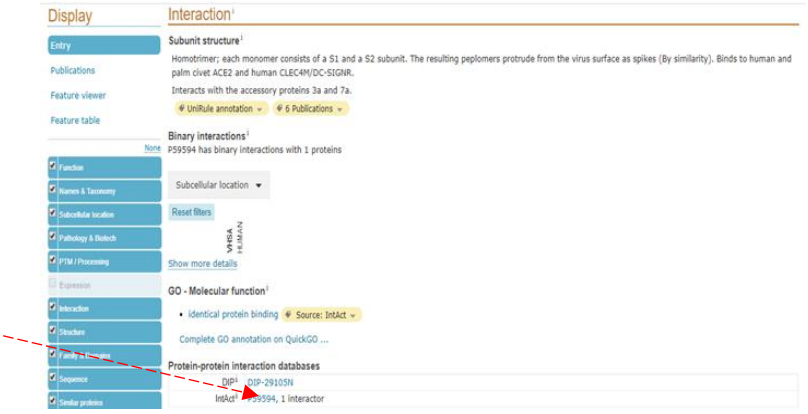

Figure 3 Interaction information of spike glycoprotein

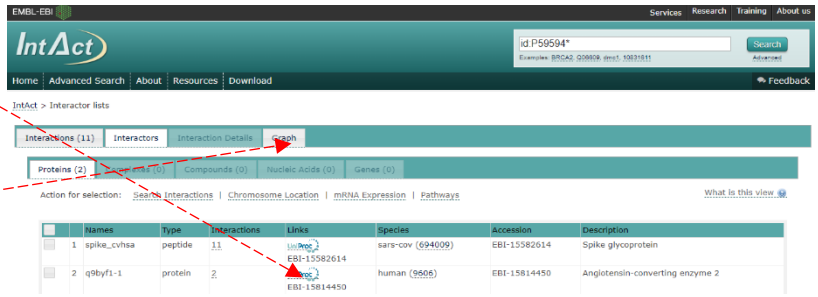

Figure 4 Proteins which can interact with spike glycoprotein

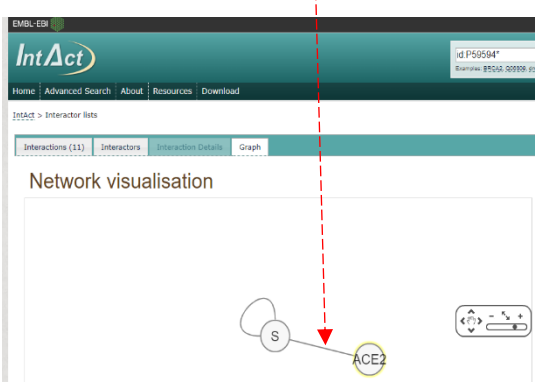

Figure 6 Network visualization of two proteins

Figure 5 Information of Angiotensin-converting enzyme 2

Result:
UniProtKB reports proteins(Angiotensin-converting enzyme 2) that can interact with and bind to spike glycoprotein of human SARS coronavirus. Human angiotensin-converting enzyme 2 (ACE2) is a functional receptor for SARS coronavirus (SARS-CoV).

iii) Critique of UniProtKB

UniProtKB follows the FAIR (Findable, Accessible, Interoperable, and Reusable) principles. Researchers will be able to perform queries and computational analyses across datasets. Suggestions to improve UniProKB include:

1). Besides collects and centralizes functional information on proteins across a wide range of species, UniProtKB can also include links to other external databases. For example, the latest research progress of diseases, clinical and animal experimental reports.

2). Add disease-related information. For example, no available disease data for the severe acute respiratory syndrome, despite Severe acute respiratory syndrome (SARS) is a viral respiratory illness caused by a coronavirus called SARS-associated coronavirus (SARS-CoV).