

Cross-linking based protein network web tutorial

This cross-linking dataset comprises a repository of close-proximity cross-link based protein interactions from mouse hippocampus and cerebellum. To facilitate its accessibility, we developed an interactive web portal (<http://xlink.cncr.nl>). This tutorial demonstrates the use of this cross-linking platform and provides an example for the interpretation of the data.

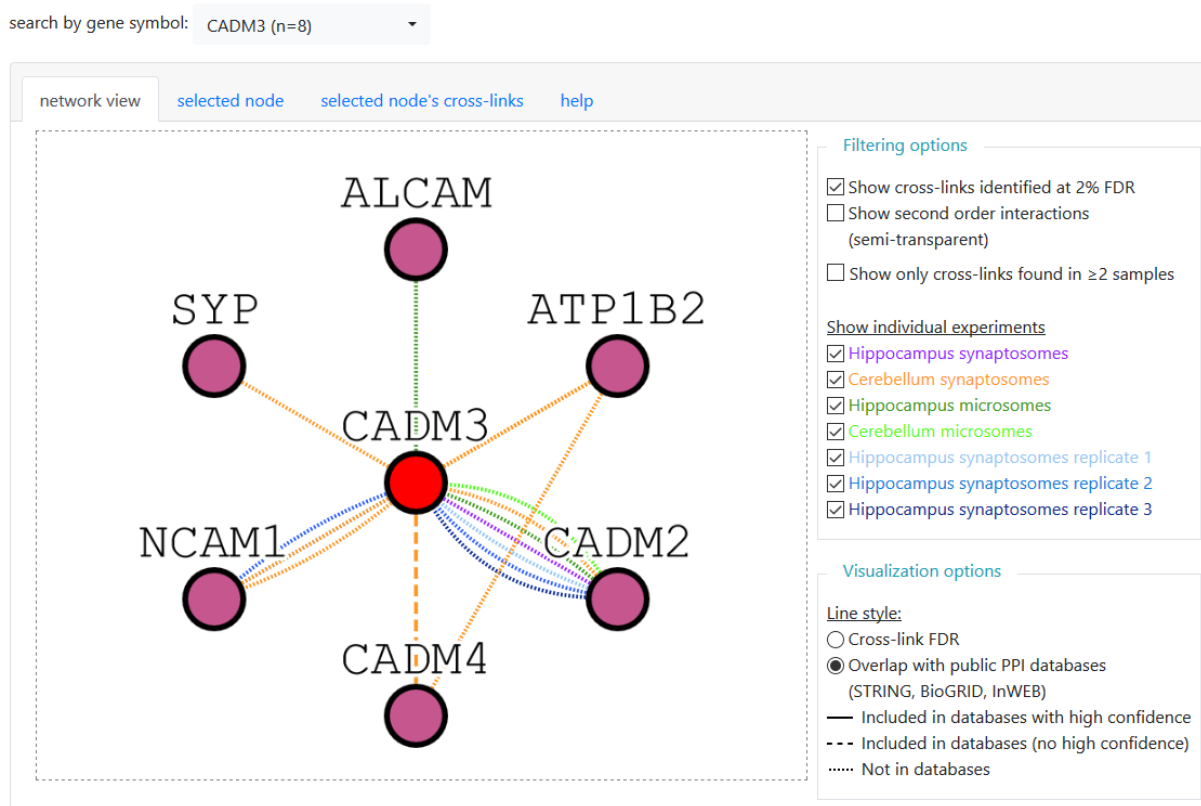
A complete description of the synapse cross-linking data can be found here: Gonzalez-Lozano, M.A. *et al.* Stitching the synapse: cross-linking mass spectrometry to resolve synaptic protein complexes. *Sci Adv.* **X**, XXXXX (2019).

Exploring the network

Go to <http://xlink.cncr.nl>. On this page, you will find the most relevant information regarding the dataset, the basic commands for the network and download links.

Scroll down to the network view panel:

1. Click a node to select it or use the search box to find your protein of interest (labeled by official gene symbols). Use the mouse scroll wheel to zoom in/out. In this example we searched for Cadm3.

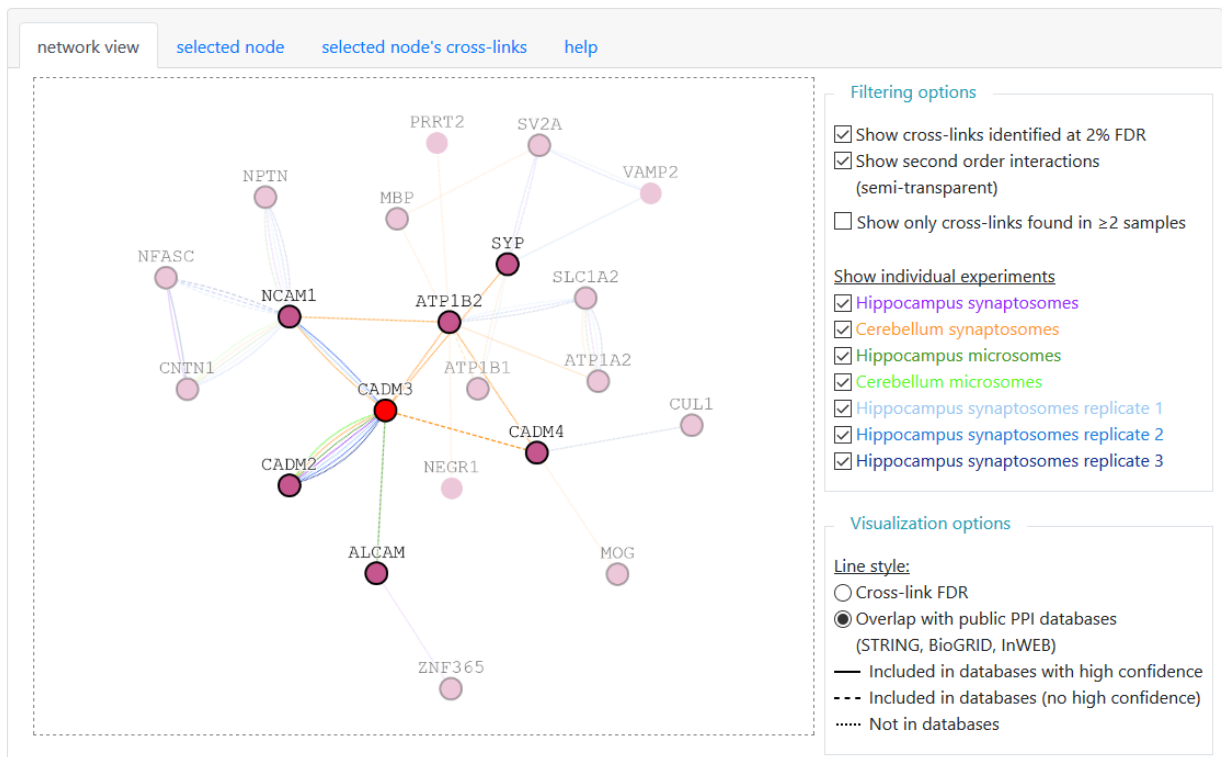


Each node represents a protein and the edges represent the protein-protein interaction as identified by cross-linking. Black borders indicate the identification of intraprotein cross-links. Edge color

indicates the individual experiment in which the cross-links were identified. Edge style can be used to visualize the overlap with public protein-protein interaction databases (STRINGDB, InWEB and/or BioGRID) and identification confidence (FDR). The number in the search box indicates the number of protein cross-linked to the protein selected.

- Only nodes directly connected to the selected protein are displayed in the network view. To extend the network, select the select *Show second order interactions* under the filtering options panel. The additional shell of interactions is visualized semi-transparent.

search by gene symbol: CADM3 (n=8) ▾



- Additional filtering options are available to visualize only high confidence cross-links (1% FDR), cross-link found in two or more samples and/or in a specific selection of experiments (e.g. cross-links found in a certain brain area or subcellular fraction).
- Nodes can be moved to improve visualization by dragging the node (click and hold-down the mouse button on any node, then drag). Click the currently selected protein (in red) to reset the network layout.
- In the case of Cadm3, we see a set of proteins identified in different number of samples with high confidence (1% FDR), which interactions are not described in public databases (dotted line edges).

Protein information

Click 'selected node' tab to inspect the information of the selected protein. It includes protein name, protein length (number of amino acids), number of lysines, residue ranges for known protein domains and links to several resources, including several publicly accessible protein interaction databases.

search by gene symbol:

| network view | selected node | selected node's cross-links | help |
|------------------------------|--|---|----------------------|
| Gene Symbol | CADM3 | | |
| Name | Cell adhesion molecule 3 (Immunoglobulin superfamily member 4B) (IgSF4B) (Nectin-like protein 1) (NECL-1) (Synaptic cell adhesion molecule 3) (TSLC1-like protein 1) | | |
| Uniprot ID | Q99N28 | | |
| Protein Length | 396 | | |
| Lysine Counts | 19 | | |
| Uniprot Domain info | DOMAIN 23 124 Ig-like V-type.; DOMAIN 128 226 Ig-like C2-type 1.; DOMAIN 231 313 Ig-like C2-type 2. | | |
| MGI mouse DB | MGI:2137858 | | |
| HGNC ortholog | HGNC:17601 | | |
| StringDB | ENSMUSP00000106851 | | |
| bioGRID | 220505 | | |
| inweb | ID: uniprotkb:Q8N126 manual search on inweb website | | |

In our example, we observe that Cadm3 (Cell adhesion molecule 3) is a protein of 396 residues with 3 domains Ig-like domains: V-type between residues 23-124, C2-type 1 between residues 128-226 and C2-type 2 between residues 231-313. For more information you can click the links to UniProt, MGI or HGNC.

Cross-linking interaction information

Click 'selected node's cross-links' tab to inspect the information of the interactions shown in the network view panel. A table with several columns will display:

- PPI: indicates the protein-protein interactions, including within (intraprotein cross-links) and between proteins (interprotein cross-links).
- Samples (value 0-7): number of samples in which the interaction has been found. Seven individual experiments were performed in this study.
- Unique xlink position: number of different unique Lys-Lys position cross-links.
- Intraprotein (true/false): indicates intraprotein (true) or interprotein (false) interactions.
- Also in PPI databases (value 0-3): indicate the number of public protein interaction databases in which the connection is found (STRINGDB, InWEB and/or BioGRID).

- High confidence in public databases (true/false): indicate if the interaction has been found with high confidence in any public database (see below).
- Public database score (low/medium/high): confidence of the interaction reported in the public databases (see confidence criteria below). Hover over the text to examine the evidence details, including the score, assay and PMID of the reporting publication when available.

search by gene symbol:

network view selected node selected node's cross-links help

Show entries Search:

| | PPI ^{↑↓} | samples ^{↑↓} | unique xlink positions ^{↑↓} | intraprotein ^{↑↓} | Also in n PPI databases ^{↑↓} | High confidence in StringDB, bioGRID or inweb ^{↑↓} | StringDB highscore (hover for details) ^{↑↓} | bioGRID highscore (hover for details) ^{↑↓} | inweb highscore (hover for details) |
|--|-------------------|-----------------------|--------------------------------------|----------------------------|---------------------------------------|---|--|---|-------------------------------------|
| | ALCAM*ALCAM | 6 | 3 | true | 0 | false | | | |
| | ALCAM*CADM3 | 1 | 1 | false | 0 | false | | | |
| | ATP1B2*ATP1B2 | 7 | 1 | true | 0 | false | | | |
| | ATP1B2*CADM3 | 1 | 1 | false | 0 | false | | | |
| | ATP1B2*CADM4 | 1 | 1 | false | 0 | false | | | |
| | ATP1B2*NCAM1 | 1 | 2 | false | 0 | false | | | |
| | CADM2*CADM2 | 7 | 5 | true | 0 | false | | | |
| | CADM2*CADM3 | 7 | 2 | false | 0 | false | | | |
| | CADM3*CADM3 | 7 | 3 | true | 0 | false | | | |
| | CADM3*CADM4 | 1 | 1 | false | 2 | false | medium | | low |

Showing 1 to 10 of 15 entries Previous 1 2 Next

Click the green '+' symbol on the left column to expand the cross-linking information for the selected interaction. Use the search box on the top right corner to easily find the interaction of interest.

- Xlink id: cross-link identifier.
- Sample: indicates the experiment in which the cross-link was identified. Brain region and subcellular fraction are indicated as: Cb, cerebellum; Hp, hippocampus; Syn, synaptosomes; M, microsomes.
- FDR cutoff (value 1-2): provides false discovery rate of the identification. An FDR of 1 is considered of high confidence.
- From and To domain: indicates the protein domain in which the cross-link was identified.

- From and To xlink site: indicates the UniProt ID, the sequence identified and the position of the residue cross-linked. Notice that the direction of the interaction with this type of data is not relevant.

search by gene symbol:

network view selected node selected node's cross-links help

Show 10 entries Search: ncam

| | PPI | samples | unique xlink positions | intraprotein | Also in n PPI databases | High confidence in StringDB, bioGRID or inweb | StringDB highscore (hover for details) | bioGRID highscore (hover for details) | inweb highscore (hover for details) |
|----------------------------------|--------------|---------|------------------------|--------------|-------------------------|---|--|---------------------------------------|-------------------------------------|
| <input checked="" type="radio"/> | ATP1B2*NCAM1 | 1 | 2 | false | 0 | false | | | |
| <input type="radio"/> | CADM3*NCAM1 | 3 | 1 | false | 0 | false | | | |

| xlink id | sample | FDR cutoff | from domain | to domain | from xlink site | to xlink site |
|----------|-------------|------------|-------------------|------------------------|----------------------------|--------------------|
| 4383 | cb_syn | 1 | Ig-like C2-type 2 | Fibronectin type-III 2 | Q99N28;EGQKLLHCEGR;244 | P13595;HYLVKYN;634 |
| 11030 | hp_syn_rep1 | 1 | Ig-like C2-type 1 | Fibronectin type-III 2 | Q99N28;EKETATLNCQSSGSK;142 | P13595;HYLVKYN;634 |
| 11030 | hp_syn_rep2 | 1 | Ig-like C2-type 1 | Fibronectin type-III 2 | Q99N28;EKETATLNCQSSGSK;142 | P13595;HYLVKYN;634 |

Analysis and interpretation of the results

Following the previous example, Cadm3 was found cross-linked to several proteins, which have not been reported in the public databases, such as Ncam1 and Alcam. Both have been identified cross-linked to the residue 244 of Cadm3 with high confidence (=FDR 1), which corresponds to the Ig-like C2-type 2 domain of Cadm3, which may suggest a similar binding mechanism. In addition, Ncam1 was also found cross-linked to the Ig-like C2-type 1 domain of Cadm3 in two independent experiments, which increases the confidence of the identification. The cross-links with Ncam1 were found in three synaptosomal samples from hippocampus and cerebellum, while the cross-link with Alcam was found only in hippocampus microsomes. If we now search for Ncam1 and Alcam, we notice that both proteins have identified cross-links to several proteins across the different brain regions and subcellular fractions. Since all three protein were identified in every condition, it seems unlikely that a differential expression or localization of these proteins is the main factor driving the sample specificity of the cross-links between Ncam1 and Alcam with Cadm3. Based on these data, we can hypothesize that Cadm3 may interact with Ncam1 and Alcam in a similar manner (same domain family) and may be differentially regulated in different brain areas and subcellular fractions. Following the same strategy, our cross-linking dataset could be used to select candidates for follow-up experiments and to infer interaction mechanisms, binding site competition, etc. For more protein interaction information involving the protein of interest, not present

in our cross-linking dataset, we can click the links to StringDB, InWEB or BioGRID in the 'selected node' tab.

Protein-protein interaction databases confidence scoring

Literature protein interactions were classified in confidence levels (low, medium and high) depending on the data curation of the different sources:

- STRING: only experimentally determined interactions were considered. Interaction with a combined score < 400 were considered as low, 400-700 as medium and ≥ 700 as high confidence.
- InWEB: interactions with a final score < 0.2 or no score provided were considered as low, 0.2-0.7 as medium and ≥ 0.7 as high confidence.
- BioGRID: interactions evidenced by low throughput experiments were considered as high and the rest as low confidence.

Final remark

In this tutorial we have shown how to explore our extensive cross-linking collection for the generation of hypothesis on novel protein-protein interactions and protein binding regions. To retrieve the most reliable protein interactions for follow-up studies, we recommend to focus on high confidence cross-links (1% FDR) identified in several samples. The complete dataset can alternatively be found as Table S1 in the original publication.

This data can be further exploited in combination with structural resources as described in:

Gonzalez-Lozano, M.A. *et al.* Stitching the synapse: cross-linking mass spectrometry to resolve synaptic protein complexes. *Sci Adv.* **X**, XXXXX (2019).