

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- | | | |
|-------------------------------------|-------------------------------------|--|
| <input type="checkbox"/> | <input checked="" type="checkbox"/> | The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> | A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> | The statistical test(s) used AND whether they are one- or two-sided
<i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i> |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> | A description of all covariates tested |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> | A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> | A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals) |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> | For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
<i>Give P values as exact values whenever suitable.</i> |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> | For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> | For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> | Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated |

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection

Mass spectrometry data were acquired with an Q Exactive HF mass spectrometer (Thermo Fisher Scientific, Bremen, Germany, Tune 2.9.3.2948, Xcalibur 4.1.31, SII for Xcalibur 1.4) coupled to an Ultimate 3000 RSLC nano system (Dionex, Thermo Fisher Scientific, Sunnyvale, USA).

Data analysis

xiSEARCH (v. 1.6.753), xiFDR (v. 2.1.3 and 2.1.5), Comet (v. 2019010), msConvert (3.0.20175.cbf82d022), xiRT (v. >1.1), pLink2 (v. 2.3.9), seaborn (0.11.0), SHAP (0.36.0), sklearn (0.24.1), tensorflow (1.15 and >2), python 3.8.3/3.7, xiVIEW (https://xiview.org/xiNET_website/index.php, beta), xiRT (<https://github.com/Rappsilber-Laboratory/xiRT>)

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium (<http://proteomecentral.proteomexchange.org>) via the jPOST partner repository69 with the dataset identifier PXD020407 and DOI 10.6019/PXD020407 [<http://proteomecentral.proteomexchange.org/cgi/GetDataset?ID=PX020407>]. Raw data of the FA-Complex is available via the previously published PRIDE identifier PXD014282 [<http://proteomecentral.proteomexchange.org/cgi/>

GetDataset?ID=PXD014282]. Additional files and intermediate results are available via Zenodo 10.5281/zenodo.4270324 [https://zenodo.org/record/4270324]. Source data are provided with this paper. PPI data was retrieved from STRING data (https://string-db.org/, v11) and APID (http://cicblade.dep.usal.es:8080/APID/init.action, downloaded 09/2019).

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences Behavioural & social sciences Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Sample size were not predetermined. Our rationale was to choose a sample size large enough to generate a diverse collection of cross-linked peptides from many protein-protein-interactions to train a deep neural network. As shown in Fig. 2 (d,e,f,g) the collected data suffices to train accurate deep learning models.
Data exclusions	No data were excluded.
Replication	Most samples were measured once because replicate measurements would have tremendously increased the acquisition time and cost. However, note that we also successfully tested our workflow on a publicly available data set.
Randomization	Samples were acquired in a non-random order to minimize the influence of sample carry-over within the acquisition queue and so to improve any prediction from machine learning.
Blinding	Not applicable, because there is no expected observer bias.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involvement in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology and archaeology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Human research participants
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern

Methods

n/a	Involvement in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging