Supplementary Materials for

Protein Interaction Landscapes Revealed by Advanced *In Vivo* Cross-linking Mass Spectrometry

Andrew Wheat[1], Clinton Yu[1], Xiaorong Wang[1], Anthony Burke[2], Ilan Chemmama[3], Robyn M. Kaake[4], Peter Baker[5], Scott D. Rychnovsky[2], Jing Yang[6], Lan Huang[1*]

[1]Department of Physiology & Biophysics, University of California, Irvine, CA 92697, USA

[2]Department of Chemistry, University of California, Irvine, CA 92697, USA

[3]Department of Cellular and Molecular Pharmacology, University of California, San Francisco, CA, 94158, USA

[4]Gladstone Institutes, University of California, San Francisco, CA, 94158, USA

[5]Department of Pharmaceutical Chemistry, University of California, San Francisco, CA, 94158, USA

[6]State Key Laboratory of Proteomics, Beijing Proteome Research Center, Beijing 102206, China

*Correspondence should be addressed to Dr. Lan Huang (lanhuang@uci.edu)

Medical Science I, D233

Department of Physiology & Biophysics

University of California, Irvine

Irvine, CA 92697-4560

Phone: (949) 824-8548

Fax: (949) 824-8540

SUPPLEMENTARY METHODS

Immunoblotting Analysis

Cross-linking efficiency was evaluated by SDS-PAGE and immunoblotting with streptavidin-HRP conjugate probing Rpn11-HTBH. Protein extraction efficiency using lysis buffers at different pH was evaluated by probing Rpn11-HTBH and H2B using their specific antibodies.

Separation of Cross-linked Peptides by SEC

Peptide separation by SEC was performed similarly as described (1). Briefly, dried peptides were reconstituted in SEC mobile phase (0.1% formic acid and 30% ACN) and separated on a Superdex Peptide PC 3.2/30 column (300 x 3.2 mm) at a flow rate of 50 µL/min, monitored at 215, 254 and 280 nm UV absorbance simultaneously. Two minute fractions were collected, and only fraction 24 and 26 were subjected to LC MS^n analyses.

In Vitro Cross-linking of BSA

For benchmarking of cross-link enrichment, standard protein BSA was cross-linked *in vitro* by Alkyne-A-DSBSO similarly as described (2). 50 µM BSA in PBS buffer (pH 7.4) was reacted with Alkyne-A-DSBSO at a molar ratio of 1:10 at room temperature for 1 hr. The reaction was quenched by adding 1 M NH_4HCO_3 solution to a final concentration of 50 mM for 15 min at room temperature. Cross-linked BSA was digested using the FASP protocol (3).

Digestion of Cross-linked Proteins

Cross-linked proteins were digested using a modified FASP protocol with 30,000 NMWL Microcon centrifugal tubes (3). Briefly, proteins were reduced with 2 mM TCEP for 30 min at room temperature, followed by alkylation with 10 mM chloroacetamide for 30 min in the dark at room temperature. Samples were then reconstituted in 8M urea, 25 mM NH_4HCO_3 buffer and

digested first with 0.75% (w/w) Lys-C at 37˚C for 4 hrs. Next, the urea concentration was reduced to 1.5 M and samples were further digested with trypsin (1.5% (w/w)) at 37˚C overnight. The peptide digests were then collected and desalted using a Waters Sep-Pak C18 cartridge prior to cross-link enrichment and MS analysis.

Benchmarking of Cross-link Enrichment

To evaluate the sensitivity, specificity and efficiency of click-chemistry based enrichment, we have utilized Alkyne-A-DSBSO cross-linked BSA and $293^{Rpn11\text{-}HTBH}$ cell lysates. For the first comparative analysis, sensitivity was evaluated by spiking 1 µg of Alkyne-A-DSBSO cross-linked peptides of BSA into increasing amounts of non-cross-linked HEK 293 cell lysate digests (1, 10, or 100 µg) respectively. Recovery was evaluated by spiking decreasing amounts of BSA peptides (100, 10, or 1µg) into a 100 µg of non-cross-linked HEK 293 cell lysate digests. The complex mixtures were subjected to click-chemistry based enrichment as described above. For the second comparative analysis, Alkyne-A-DSBSO cross-linked $293^{Rpn11\text{-}HTBH}$ cell lysates were analyzed by LC MS/MS and LC MS^n using either enriched or non-enriched samples. The cross-link enrichment was done as described above and the same amount of samples were analyzed. MS/MS was used to identify non-cross-linked peptide and MS^3 analysis was used to identify cross-linked peptides. Foe each comparison, equal amounts of the samples were injected for MS analysis. Efficiency was calculated by dividing the ratio of the number of cross-linked peptides to the total number of all peptides in the elution after enrichment by ratio of the number of cross-linked peptides to the total number of all peptides in the non-enriched sample. Specificity was calculated by dividing the number of cross-linked peptides by the total number of all peptides in the elution (4).

LC MS/MS Analysis

For benchmarking of cross-link enrichment, LC MS/MS analysis was also carried out to identify non-cross-linked peptides using an Orbitrap Fusion Lumos (Thermo Scientific) coupled on-line with an Ultimate 3000 HPLC (Thermo Scientific). Each cycle consisted of one full FT scan mass spectrum (375 – 1500 m/z, resolution of 60,000 at m/z 400) followed by data-dependent MS/MS acquired in the linear ion trap with collision energy of 30% at top speed for 3 s.

LC MS^n Analysis of Cross-linked Peptides

LC MS^n analyses were performed using an Orbitrap Fusion Lumos (Thermo Scientific) coupled on-line with an Ultimate 3000 HPLC (Thermo Scientific) as previously described (5). MS^1 scans were measured in the Orbitrap with a scan range from 375 to 1500 *m/z*, resolution set to 120,000, and the AGC target set to 4×10^5. MS^1 acquisition was performed in top speed mode with a cycle time of 5 s. For MS^3 analysis, either 3+ or 4+ and up charged ions were selected for MS^2. For MS^3, 2 types of acquisition methods were employed: 1) top4 CID-MS^3 acquisition; 2) targeted CID-MS^3 acquisition with the mass difference (Δ=182.0071, $C_5H_{10}O_3S_2$) between characteristic ion pairs of DSBSO cross-linked peptides in MS^2 ($\Delta= \alpha_T-\alpha_A=\beta_T-\beta_A$). MS^1 and MS^2 scans were acquired in the Orbitrap whereas MS^3 scans were detected in the ion trap. For MS^2 scans, the resolution was set to 30,000, the AGC target 5e4, the precursor isolation width was 1.6 *m/z*, and the maximum injection time was 100 ms for CID. The CID-MS^2 normalized collision energy was 25%. For MS^3 scans, CID was used with a collision energy of 35%, the AGC target was set to 2×10^4, and the maximum injection time was set to 120 ms. Raw data were deposited in the PRIDE repository with the identifier PXD012788 (reviewer76771@ebi.ac.uk and password: wdbSfDb6).

Database Searching and Cross-link Identification

Raw LC MS/MS and LC MS^n data were converted to MGF files using ProteoWizard MSConvert (v. 3.0.10738) respectively. Extracted MS^3 and MS/MS spectra were subjected to protein database searching via Batch-Tag within a developmental version of Protein Prospector (v. 5.19.1, University of California, San Francisco) against a randomly concatenated decoy-containing *Homo sapien* database (SwissProt.2017.11.01.random.concat; 20,240 entries). The mass tolerances for parent ions were set to ±20 ppm and fragment ions were set to 0.6 Da. Trypsin was set as the enzyme with three maximum missed cleavages for MS^3 data and two maximum missed cleavages for MS/MS data. Carbamidomethylation was set as a constant modification for cysteine, while protein N-terminal acetylation, N-terminal conversion of glutamine to pyroglutamic acid, asparagine deamidation, and methionine oxidation were selected as variable modifications. For MS^3 analysis, three additional modifications were added for uncleaved lysines and protein N-termini: alkene (C_3H_2O, +54.0106 Da), unsaturated thiol ($C_8H_{12}S_2O_4$, +236.0177 Da), sulfenic acid ($C_8H_{14}S_2O_5$, +254.0283 Da), corresponding to remnant moieties of DSBSO after cross-link cleavage. Due to the tendency of the DSBSO thiol moiety to undergo additional cleavage into the alkene moiety alongside backbone fragmentation during MS^3 analysis, we have incorporated such neutral loss in Batch-tag to facilitate the identification of thiol-modified peptides during database searching using Protein prospector (6). The in-house software XL-Tools was used to automatically identify, summarize and validate cross-linked peptides based on Protein Prospector database search results and MS^n data (7). For MS/MS data, peptides were identified at 5% FDR for benchmarking analysis. For all MS^3 analyses, the FDR of cross-linked peptide identifications was determined to be < 1%, based on previous reports (7).

Replicates and Statistics

Two biological replicates were acquired for both extraction methods, i.e. pH 8-only and sequential pH (pH 7 + 8) extractions. The latter extraction resulted in two samples. For each biological replicate, two SEC fractions were subjected to MS analysis. For each fraction, four types of LC MS^n analyses were acquired.

SUPPLEMENTARY RESULTS

XL-PPIs in the UPS Network

In this work, a new interaction was identified between UBE2L3 and TBCB (UBE2L3:K64 and K67 to TBCB:K188). Since TBCB is a known substrate of E3 ligase gigaxonin and cullin-3 that requires UBE2L3 as the E2 ubiquitin-conjugating enzyme during protein ubiquitination, the UBE2L3-TBCB cross-link likely represents an enzyme-substrate interaction. Similarly, the interaction between RNF20 and H2B (RNF20: K21 and H2B: K121) also describes an enzyme-substrate interaction, as RNF20 is a known E3 ligase of H2B that is important for tumor suppression (8). Apart from ubiquitination substrates, the identified interactions may have other functional roles in ubiquitination-related biological processes. One interesting interaction identified here is between HUWE1 and PARP, a critical DNA repair protein frequently upregulated in breast cancer. Although a recent study has identified RNF144A as a PARP-targeting E3, our data implicates a novel association between HUWE1 and PARP. However, whether HUWE1 acts as an E3 for PARP or if the two proteins associate to form a complex underlying ubiquitination of other substrates remains to be determined. Interestingly, PARP was also found to interact with MRPS9, a 28S mitoribosomal protein component involved in the DNA damage response and detection of DNA damage, providing a potential link for the role of mitochondrial health in breast cancer growth and progression. As a result, HUWE1—known to be

upregulated in breast cancer cells—may serve as an important E3 ubiquitin ligase machinery related to tumorigenesis.

Relevance of XL-PPIs to Disease Networks

To understand the relevance of protein interactions with diseases, we correlated the identified XL-PPIs with 9,411 disease-associated human genes described in DisGeNET. In comparison, 70% of our XL-proteome were related to a large number of diseases (Dataset S8). Among them, breast cancer was well-represented by XL-proteome, which contains 17% (i.e. 84 proteins) of the known breast cancer-associated genes, described by 751 cross-linked peptides and 34 XL-PPIs. In comparison to the existing interaction databases, 90% of the identified breast-cancer related PPIs are novel. One of the identified breast cancer related interactions is between the oncogene MDM2:K344 and histone H2B:K21. MDM2, a E3 ubiquitin ligase, is the key negative regulator of the tumor suppressor p53 and can interact with histones and ubiquitinate H2B (9). However, how MDM2 interacts with H2B *in vivo* was not clear. K344 of MDM2 is near the Zinc finger motif (aa# 299-328) that is known to bind DNA and most likely assists in the interaction between MDM2 and the N-terminal tail of H2B. Given the importance of MDM2 and H2B monoubiquitination in oncogenic activities (9), their interaction may be important during tumorigenesis. One novel breast cancer associated interaction is between Y-box binding protein-1 (YBX1):K264 and Neuronal cell adhesion molecule (NRCAM):K8. It has been suspected that YBX1 downregulation decreases breast cancer cell migration and invasion, however the underlying mechanism remains elusive (10). Since NRCAM plays a role in cell communications by signaling to the actin cytoskeleton during directional cell migration, the identified interaction provides a potential link to underlie the function of YBX1 in breast cancer.

In order to understand whether cross-linked sites of XL-PPIs are in the vicinity of disease-causing mutations, we have correlated our data with ClinVar, a public archive of the relationships among disease-associated genomic variations and phenotypes. In total, 982 proteins of the XL-proteome found in the DisGeNET were also present in ClinVar database, which constitute a PPI network with 2,823 inter-protein and 695 intra-protein interactions. Among them, 19% of inter-protein and 30%% of intra-protein cross-linked lysines were within 15 AA of clinically relevant point mutations, respectively (*SI Appendix*, Fig. S10*A*). Interestingly, for the 84 XL-proteins associated with breast cancer, 32% inter-protein and 28% intra-protein lysines were found within 15 AA of a point mutation (*SI Appendix*, Fig. S10*B*). One example is the interaction between the fibroblast growth receptors FGFR1:K207 and FGFR2:K176. FGFRs are important in mediating cell proliferation, differentiation, and migration and their abnormal mutations are found in cancer cells. In particular, $FGFR2^{E163K}$ mutant is clinically relevant and this point mutation is in close proximity to the identified cross-link contact between FGFR1 and FGFR2. These regions are located at the Ig-like C2-type 2 regions of both FGFR1 and FGFR2 that are responsible for ligand binding, implying the importance of the identified FGFR1-FGFR2 interaction. Collectively, these results suggest that *in vivo* XL-MS studies may help identify clinically important PPIs to better understand mutation-dependent pathologies in the future.

REFERENCES:

1. A. Leitner *et al.*, Expanding the chemical cross-linking toolbox by the use of multiple proteases and enrichment by size exclusion chromatography. *Molecular & cellular proteomics : MCP* **11**, M111 014126 (2012).
2. C. B. Gutierrez *et al.*, Developing an Acidic Residue Reactive and Sulfoxide-Containing MS-Cleavable Homobifunctional Cross-Linker for Probing Protein-Protein Interactions. *Analytical chemistry* 10.1021/acs.analchem.6b02240 (2016).
3. J. R. Wisniewski, A. Zougman, N. Nagaraj, M. Mann, Universal sample preparation method for proteome analysis. *Nature methods* **6**, 359-362 (2009).

4. B. Steigenberger, R. J. Pieters, A. J. R. Heck, R. A. Scheltema, PhoX: An IMAC-Enrichable Cross-Linking Reagent. *ACS Cent Sci* **5**, 1514-1522 (2019).
5. C. Gutierrez *et al.*, Structural dynamics of the human COP9 signalosome revealed by cross-linking mass spectrometry and integrative modeling. *Proc Natl Acad Sci U S A* **117**, 4088-4098 (2020).
6. R. J. Chalkley, P. R. Baker, K. F. Medzihradszky, A. J. Lynn, A. L. Burlingame, In-depth analysis of tandem mass spectrometry data from disparate instrument types. *Molecular & cellular proteomics : MCP* **7**, 2386-2398 (2008).
7. X. Wang *et al.*, Molecular Details Underlying Dynamic Structures and Regulation of the Human 26S Proteasome. *Molecular & cellular proteomics : MCP* **16**, 840-854 (2017).
8. E. Shema *et al.*, The histone H2B-specific ubiquitin ligase RNF20/hBRE1 acts as a putative tumor suppressor through selective regulation of gene expression. *Genes Dev* **22**, 2664-2676 (2008).
9. N. Minsky, M. Oren, The RING domain of Mdm2 mediates histone ubiquitylation and transcriptional repression. *Mol Cell* **16**, 631-639 (2004).
10. J. P. Lim *et al.*, YBX1 gene silencing inhibits migratory and invasive potential via CORO1C in breast cancer in vitro. *BMC Cancer* **17**, 201 (2017).
11. M. Wilhelm *et al.*, Mass-spectrometry-based draft of the human proteome. *Nature* **509**, 582-587 (2014).

SUPPLEMENTARY DATASETS

Dataset S1A. The Number of BSA Cross-linked Peptides Identified for Recovery Benchmarking of Cross-link Enrichment.
Dataset S1B. The Number of BSA Cross-linked Peptides Identified for Sensitivity Benchmarking of Cross-link Enrichment.
Dataset S1C. The Number of BSA Cross-linked Peptides Identified for Sensitivity Benchmarking of Cross-link Enrichment coupled with SEC.

Dataset S2A. Detailed Summary of the Unique DSBSO Inter-protein Inter-linked Peptides Identified by MS^n.
Dataset S2B. Detailed Summary of the Unique DSBSO Intra-subunit Cross-linked peptides Identified by MS^n.
Dataset S2C. The list of the Unique Inter-protein K-K linkages Identified by MS^n.
Dataset S2D. The list of the Unique Intra-protein K-K linkages Identified by MS^n.

Dataset S3. Comparison of Cross-link Data in Proteome-wide XL-MS Studies.

Dataset S4A. Gene Ontology Biological Functions of the XL-Proteome.
Dataset S4B. Gene Ontology Cellular Compartments of the XL-Proteome.

Dataset S5. CORUM Analysis of Protein Complexes in the XL-Proteome.

Dataset S6. Distance Mapping of *In Vivo* Alkyne-A-DSBSO Cross-links onto the Known Structures of the 286 Protein Complexes.

Dataset S7. Comparison of Cross-link Reproducibility in Proteome-wide XL-MS Studies.

Dataset S8. DisGeNET Analysis of the XL-Proteome.

SUPPLEMENTARY FIGURES

Supplementary Figure 1. Evaluation of *in vivo* cross-linking efficiency by immunoblot analysis. $293^{Rpn11\text{-}HTBH}$ cells were used for cross-linking and the resulting Rpn11-containing cross-linked products were probed by Strep-HRP against HTBH tagged Rpn11, a proteasome subunit. Lane 1: uncross-linked cells used as a control; lane 2: *in vivo* Alkyne-A-DSBSO cross-linked cells.

Supplementary Figure 2. Sequential pH extraction of *in vivo* cross-linked proteins. (A) Protein extraction efficiencies at pH 6, 7 and 8 from *in vivo* Alkyne-A-DSBSO cross-linked $293^{Rpn11\text{-}HTBH}$ cells were evaluated using immunoblotting analysis. Strep-HRP was used to probe Rpn11-containing cross-linked products, and anti-H2A histone antibody was used to probe Histone 2A. Immunoblot analysis displayed reduction of cross-linked histone protein recovery for lower pH lysis buffers. **(B)** Comparison of histone-containing and non-histone cross-linked peptides identified from sequential pH (pH 7+8) and pH 8-only extracted samples based on MS^3 analysis. Striped bar represents histone-containing cross-linked peptides; solid bars non-histone cross-linked peptides. **(C)** Overlap of cross-linked proteins identified from sequential pH and pH 8-only extracted samples.

Supplementary Figure 3. Abundance distribution of the *in vivo* XL-proteome and the MS-proteome in ProteomicsDB. Protein abundance was based on iBAQ values determined by shotgun proteomics in Wilhelm, et al data(11).

Supplementary Figure 4. Cross-link reproducibility of *in vivo* Alkyne-A-DSBSO based XL-MS experiments. (A) Unique K-K overlap between biological replicates from pH 8-only extracted samples. **(B)** Unique K-K overlap between biological replicates from sequential pH (pH 7+8) extracted samples.

Supplementary Figure 5. PPI reproducibility of *in vivo* Alkyne-A-DSBSO based XL-MS experiments. PPI overlaps between the two biological replicates from **(A)** pH 8-only extracted samples and **(B)** sequential pH (pH 7+8) extracted samples. Comparison of unique K-K linkages of the shared PPIs between the two biological replicates from **(C)** pH 8-only extracted samples and **(D)** sequential pH (pH 7+8) extracted samples.

Supplementary Figure 6. The entire *in vivo* XL-PPI network of HEK 293 cells. This map consists of 2,484 nodes and 6,439 edges based on 13,904 K-K linkages.

Supplementary Figure 7. Comparisons of our *in vivo* XL-PPIs with the selected published XL-MS studies and curated PPIs. PPI overlap van diagrams among XL-PPIs determined in this study (Wheat, et al), Fasci *et al.* 2018, Chavez *et al.* 2016, and the combined PPIs from BioGRID, BioPlex, and STRING databases.

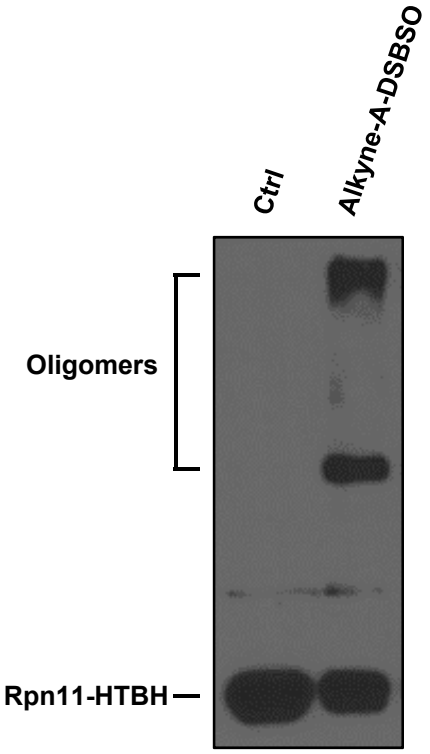
Supplementary Figure 8. Analysis of histone cross-link data. (A) Mapping of histone cross-links to a known core nucleosome structure (PDB: 5Y0D). Inter-subunit cross-links shown in red, intra-subunit cross-links shown in blue. **(B)** 2-D XL-map of the core histones and linker histone H1.3. Inter-subunit cross-links shown in red, intra-subunit cross-links shown in blue.

Supplementary Figure 9. The histone XL-PPI network. (A) Proteins identified with direct interactions with histone proteins, color-coded by associated biological functions. **(B)** 2-D XL-map among the selected proteins and histone subunits. Only inter-subunit cross-links were shown in red.

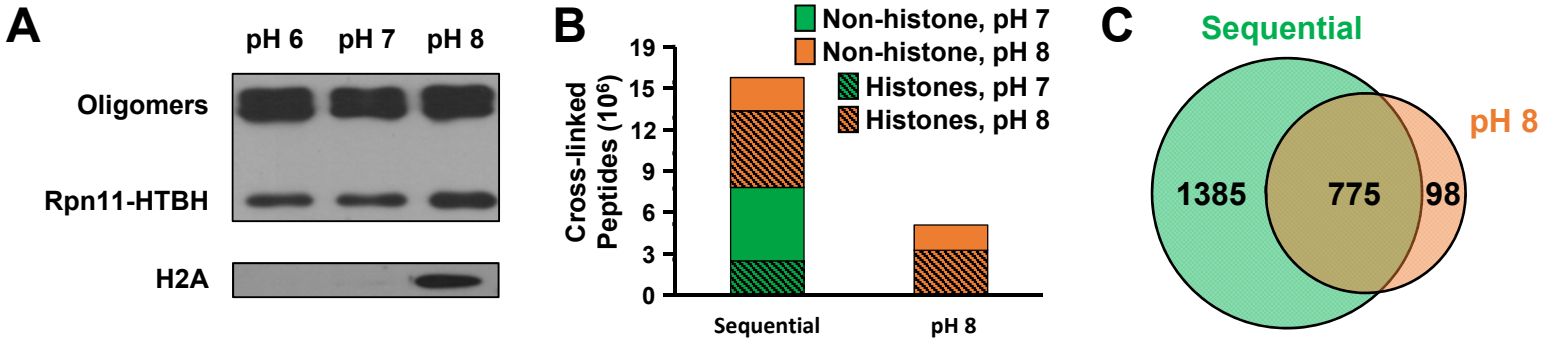
Supplementary Figure 10. Distribution of distances between cross-linked lysines and disease relevant point mutations curated by ClinVar. (A) Amino acid distance distribution of cross-linked lysines from all XL-proteome inter-protein and intra-protein interactions. **(B)** Amino acid

distance distribution of cross-linked lysines from all XL-proteome breast cancer associated inter-protein and intra-protein interaction.

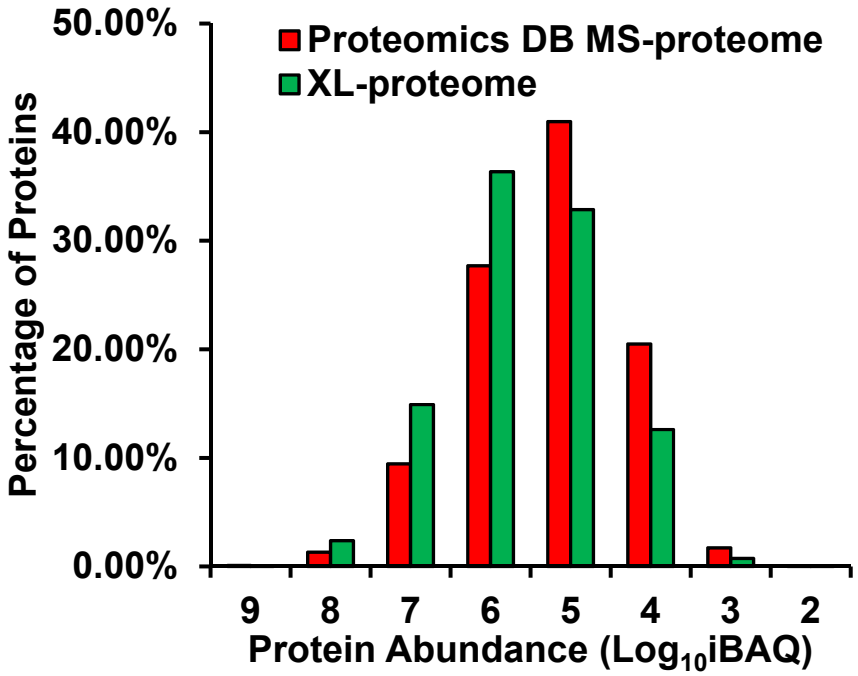
Supplementary Fig. 1



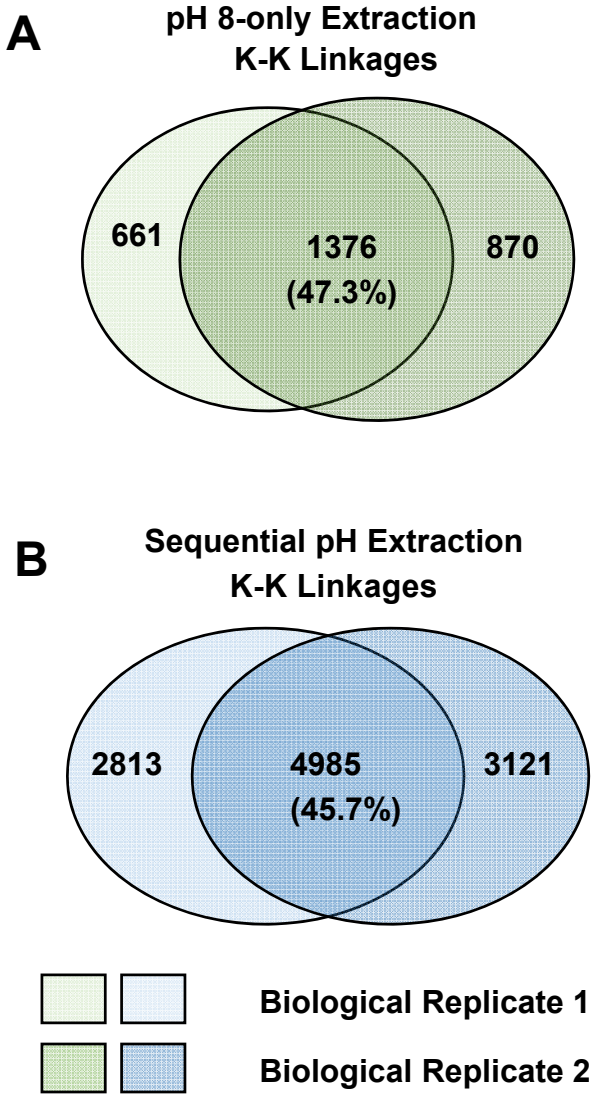
Supplementary Fig. 2



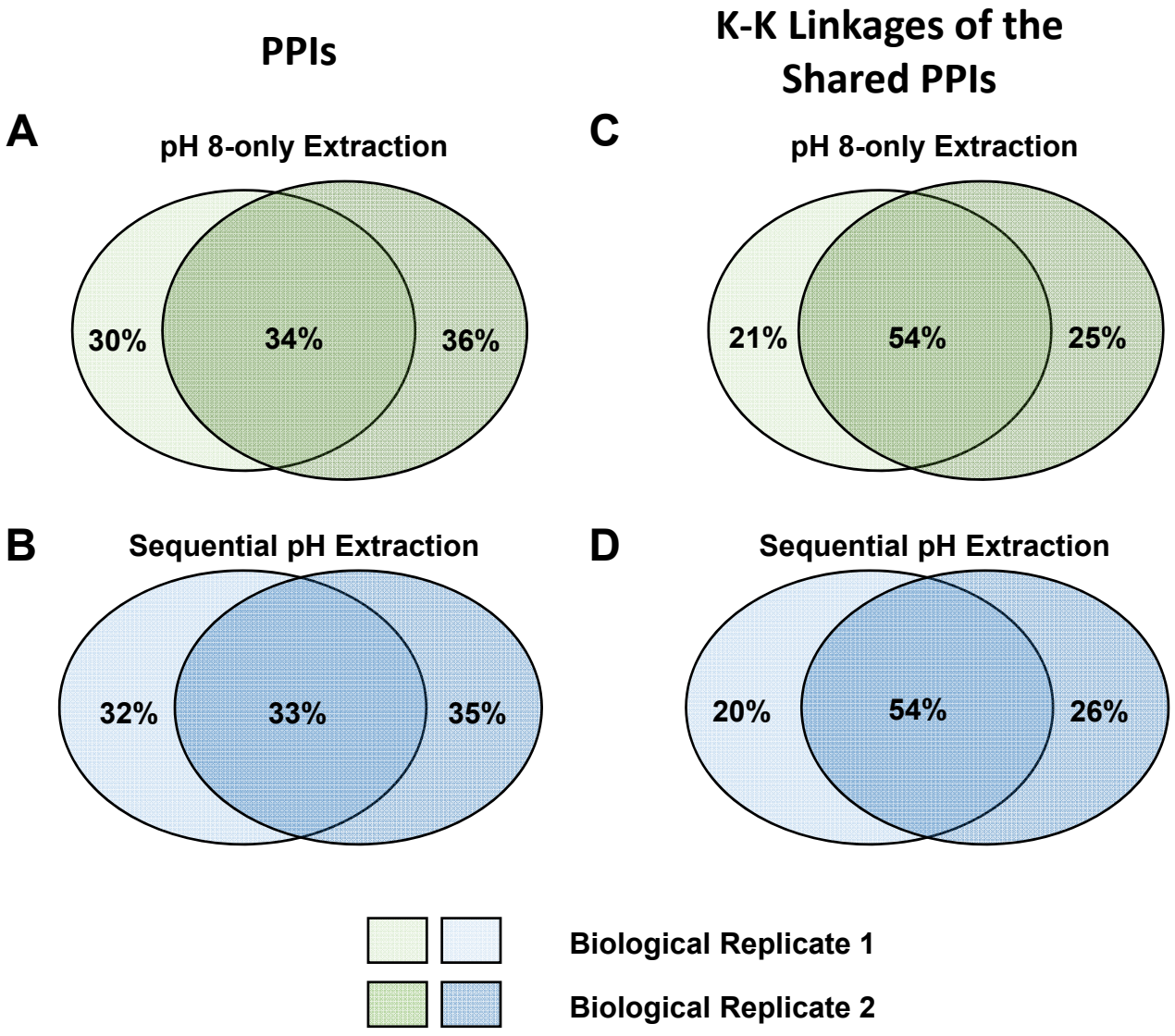
Supplementary Fig. 3



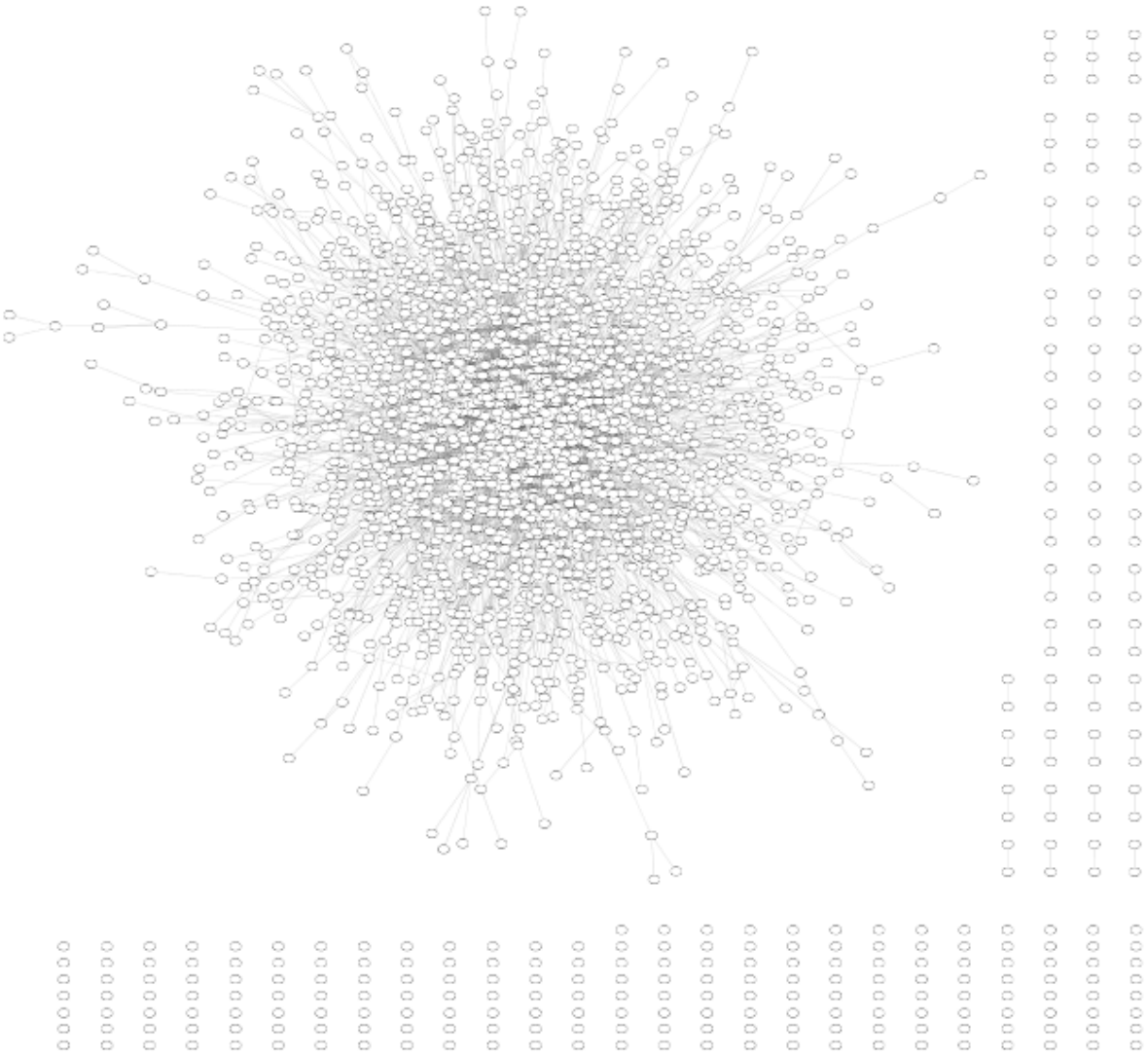
Supplementary Fig. 4



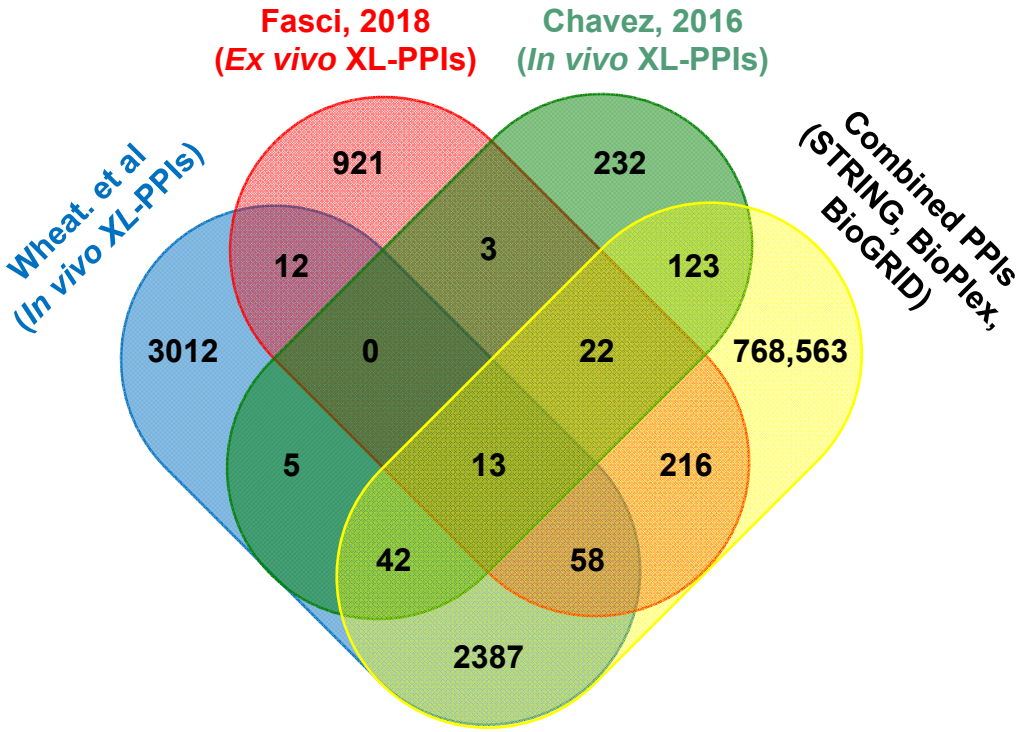
Supplementary Fig. 5



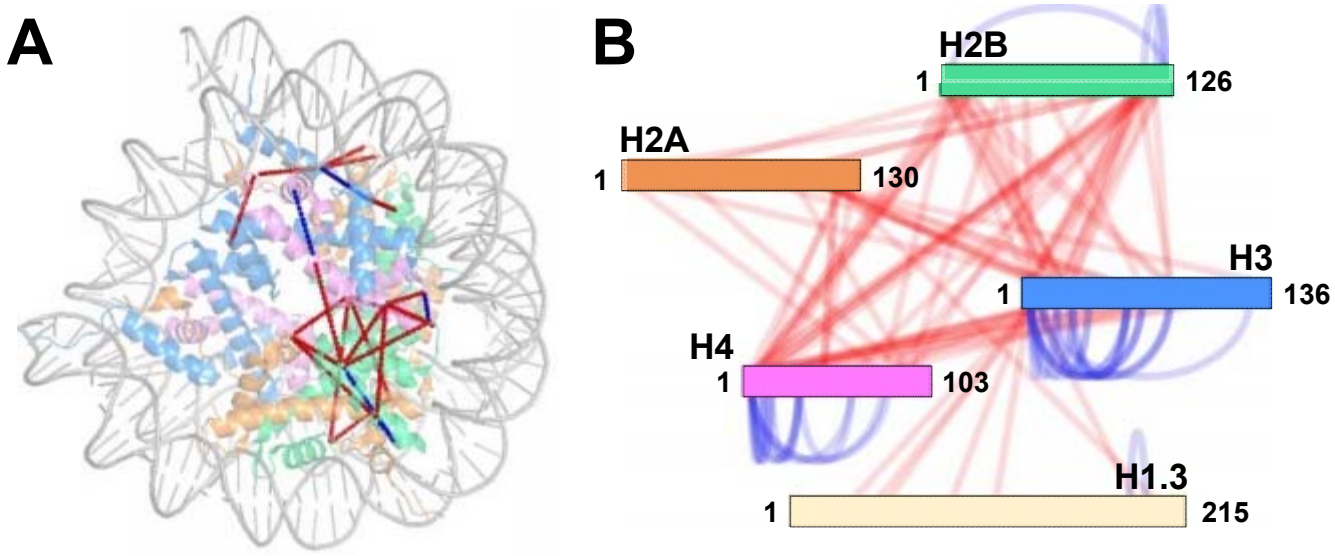
Supplementary Fig. 6



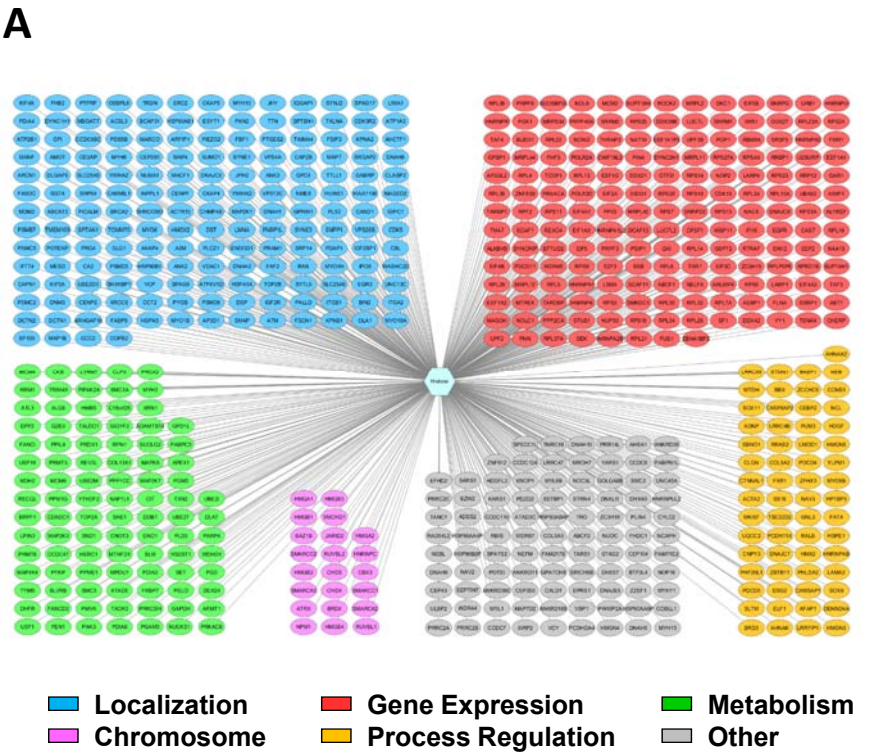
Supplementary Fig. 7



Supplementary Fig. 8



Supplementary Fig. 9



Supplementary Fig. 10