SUPPORTING INFORMATION FOR

DSBSO-based XL-MS Analysis of Breast Cancer PDX Tissues to Delineate Protein Interaction Network in Clinical Samples

Fenglong Jiao[1], Clinton Yu[1], Andrew Wheat[1], Lijun Chen[2], Tung-Shing Mamie Lih [2], Hui Zhang[2], Lan Huang[1]*

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

[2]Department of Pathology, Johns Hopkins University, Baltimore, MD 21231

*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

TABLE OF CONTENT **PAGE**

Supplemental Methods

Protein Digestion

Cross-linked proteins were transferred onto a 30 KDa FASP centrifugal filter and washed with 8 M urea. The denatured proteins were reduced with 2 mM TCEP for 30 min and alkylated with 10 mM iodoacetamide for 30 min in dark. After washing with 25 mM ammonium bicarbonate, the proteins were reconstituted in 60 µL of 8 M urea, 25 mM ammonium bicarbonate and 25 µL of 25 mM ammonium bicarbonate. Lys-C (enzyme to protein ratio of 1:100) was added to the solution and the mixture was incubated at 37°C for 4 h. Then, the concentration of urea was reduced to 1.5 M for trypsin digestion (enzyme to protein ration of 1:50) at 37°C overnight. The digested peptides were acidified by adding 1% TFA to a final concentration of 0.1% and desalted by Sep-Pak C18 cartridge. The eluents were vacuum dried and stored in -80°C before SEC fractionation.

LC-MS^n analysis

LC MS^n analysis of SEC fractions were carried out using an UltiMate 3000 RSLC coupled with an Orbitrap Fusion Lumos mass spectrometer similarly as described[1]. Samples were loaded onto a 50 cm x 75 µm Acclaim PepMap C18 column and separated over a 120 min gradient of 4% to 25% acetonitrile at a flow rate of 300 nL/min. The top 4 data-dependent MS^3 acquisition method was used for the identification of DSBSO cross-linked peptides. Ions with charge of 4^+ to 8^+ in the MS^1 scan were selected for MS^2 analysis. The top 4 abundant fragment ions in MS^2 scan were further fragmented by CID with a collision energy of 35%.

Identification of Cross-linked Peptides

The extracted MS^3 spectra were subjected to database searching using Batch-Tag within

Protein Prospector (v.6.3.5) against random concatenated human database (SwissProt. 2021. 10. 02). The mass tolerance was set as ±20 ppm for parent ions and 0.6 Da for fragment ions. Trypsin was set as the enzyme with three maximum missed cleavages allowed. Cysteine carbamidomethylation was selected as fix modification. A maximum of four variable modifications were also allowed, including methionine oxidation, N-terminal acetylation, and N-terminal conversion of glutamine to pyroglutamic acid. Three additional modifications were added for uncleaved lysines and protein N-termini: alkene (C3H2O, +54.0106 Da), unsaturated thiol (C8H12S2O4, +236.0177 Da), sulfenic acid (C8H14S2O5, +254.0283 Da), corresponding to remnant moieties of DSBSO after cross-link cleavage. 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[1,2]. The false discovery rate for cross-link identification was calculated based on a target-decoy approach as previously described[3].

TMT-based Quantitative Analysis of PDX Proteomes

Tissue lysis and downstream sample preparation for global proteomics analysis were carried out as previously described[4]. In brief, cryopulverized PDX tumor tissues were lysis in urea lysis buffer (8 M urea, 75 mM NaCl, 50 mM Tris at pH 8.0, 1 mM EDTA, 2 mg/mL aprotinin, 10 mg/mL leupeptin, 1 mM PMSF, 10 mM NaF, Phosphatase Inhibitor Cocktail 2 and Phosphatase Inhibitor Cocktail 3 [1:100 dilution], and 20mM PUGNAc). Lysates were diluted to a final concentration of 8 mg/mL with lysis buffer, and subjected to proteolytic digestion with LysC for 2h, followed by overnight trypsin digestion at RT. Tryptic peptides were desalted using Sep-Pak C18 cartridge.

Desalted peptides from P96 and P97 (two for each) were labeled using TMT16plex (Thermo Fisher Scientific). 100 µg peptides from each sample at the concentration of 2.5 µg/µL were labeled with TMT reagents in 100 mM HEPES buffer. The mass ratio of peptide to TMT is 1:2 (100 µg peptides to 200 µg TMT). After 1h labeling, the reaction was quenched by 5% Hydroxylamine for 15 min. Following labeling and quenching, peptides were mixed according to the sample-to-TMT channel mapping, concentrated and desalted and dried. The desalted, TMT-labeled samples were reconstituted in 900 µL of 20 mM ammonium formate (pH10) and 2% acetonitrile (ACN) and loaded onto a 4.6 mm x 250 mm RP Zorbax 300 A Extend-C18 column with 3.5mm size beads (Agilent). Peptides were separated at a flow-rate of 1 mL/min using an Agilent 1200 Series HPLC instrument via bHPLC with Solvent A (2% ACN, 5mM ammonium formate at pH10) and a non-linear gradient of Solvent B (90% ACN, 5mM ammonium formate at pH10). Collected fractions were concatenated into 24 fractions by combining four fractions that are 24 fractions apart (i.e., combining fractions #1, #25, #49, and #73; #2, #26, #50, and #74; and so on); a 5% aliquot of each of the 24 fractions was used for global proteomic analysis.

Global peptides (0.5 µg) were separated on an Easy nLC 1200 UPLC system (Thermo Scientific) on an in-house packed 20 cm x 75 mm diameter C18 column (1.9mm Reprosil-PurC18-AQbeads (Dr.MaischGmbH); Picofrit 10 mm opening (New Objective)). The column was heated to 50^{o}C using a column heater. The flow rate was 200 nl/min with 0.1% formic acid and 2% acetonitrile in water (A) and 0.1% formic acid, 90% acetonitrile (B). The peptides were separated with a 6%–30% B gradient in 84 min and analyzed using Thermo Fusion Lumos mass spectrometer as described[4]. All raw files were processed through MS-PyCloud[5]. In brief,

raw files were converted into mzML and searched against UniProt/Swiss-Prot protein sequence FASTA (version 20191213) containing both human and mouse proteins via MS-GF+ using the following settings: static modification of carbamidomethyl at cysteine, dynamic modifications of oxidation at methionine and TMT at lysine and N-terminus, precursor mass tolerance of 20 ppm, miss cleavages $\leq$ 2, and false discovery rate of 1% at PSM level. The protein quantification was based on the TMT report ions.

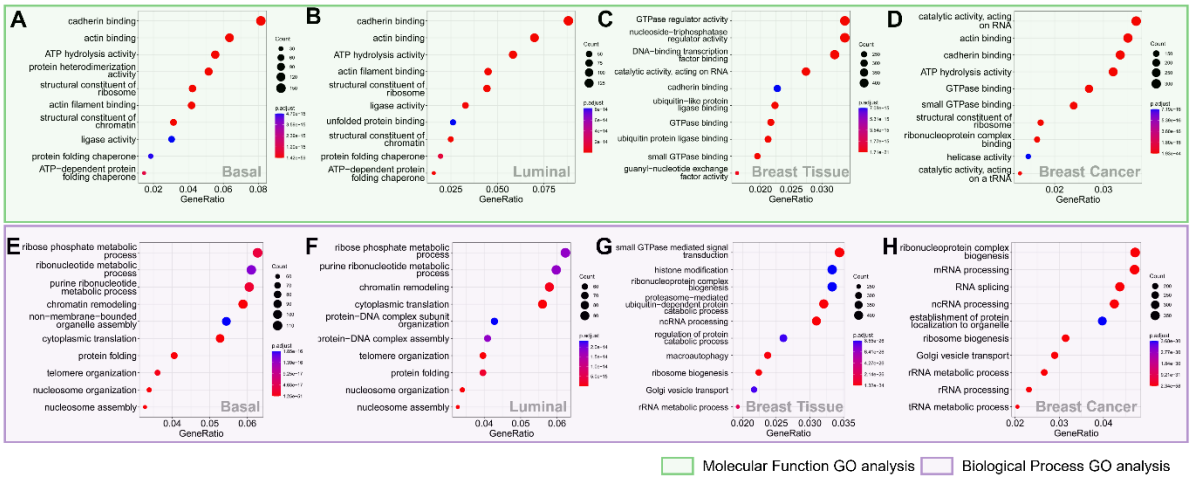

Figure S1. **GO enrichment analysis of Molecular Functions and Biological Process.** Molecular functions analysis of **(A)** XL-proteome from basal breast cancer PDX. **(B)** XL-proteome from luminal breast cancer PDX, (**C**) normal breast tissue proteome[6] and **(D)** breast cancer proteome[7].and **Biological process analysis of (E)** XL-proteome from basal breast cancer PDX. **(F)** XL-proteome from luminal breast cancer PDX, (**G**) normal breast tissue proteome[6] and **(H)** breast cancer proteome[7].

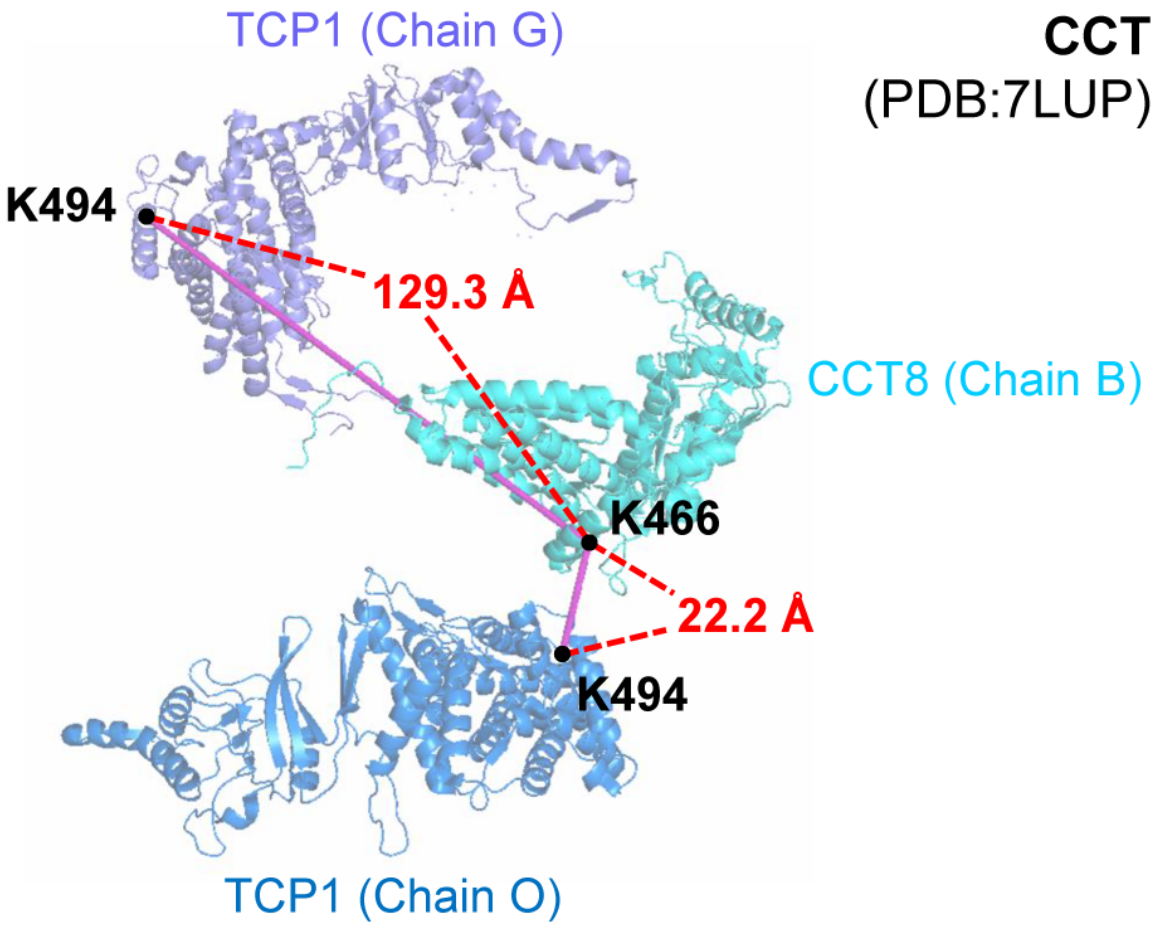

Figure S2. Distance mapping of CCT8:K466-TCP1:K494 (PDB entry: 7LUP[8]). 129.3 Å was measured from two subunits within a single ring. 22.2 Å was measured from the same two subunits on adjacent rings.

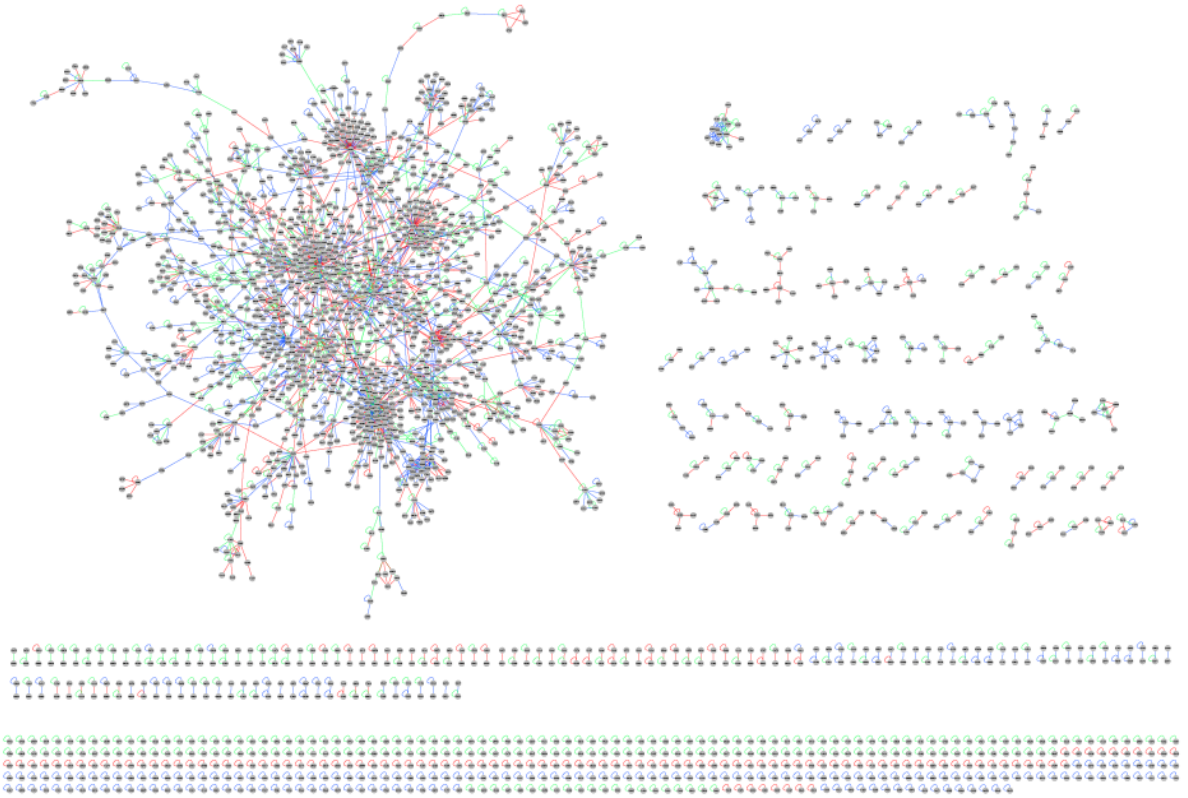

Figure S3. XL-PPI network of DSBSO cross-linked basal and luminal breast cancer PDX samples. The DSBSO XL-PPI network is composed of 2,557 nodes and 3,643 edges. Green edges: PPIs identified in basal breast cancer PDX sample only; red edges: PPIs identified in luminal breast cancer PDX sample only; grey edges: PPIs identified in both basal and luminal breast cancer PDX samples.

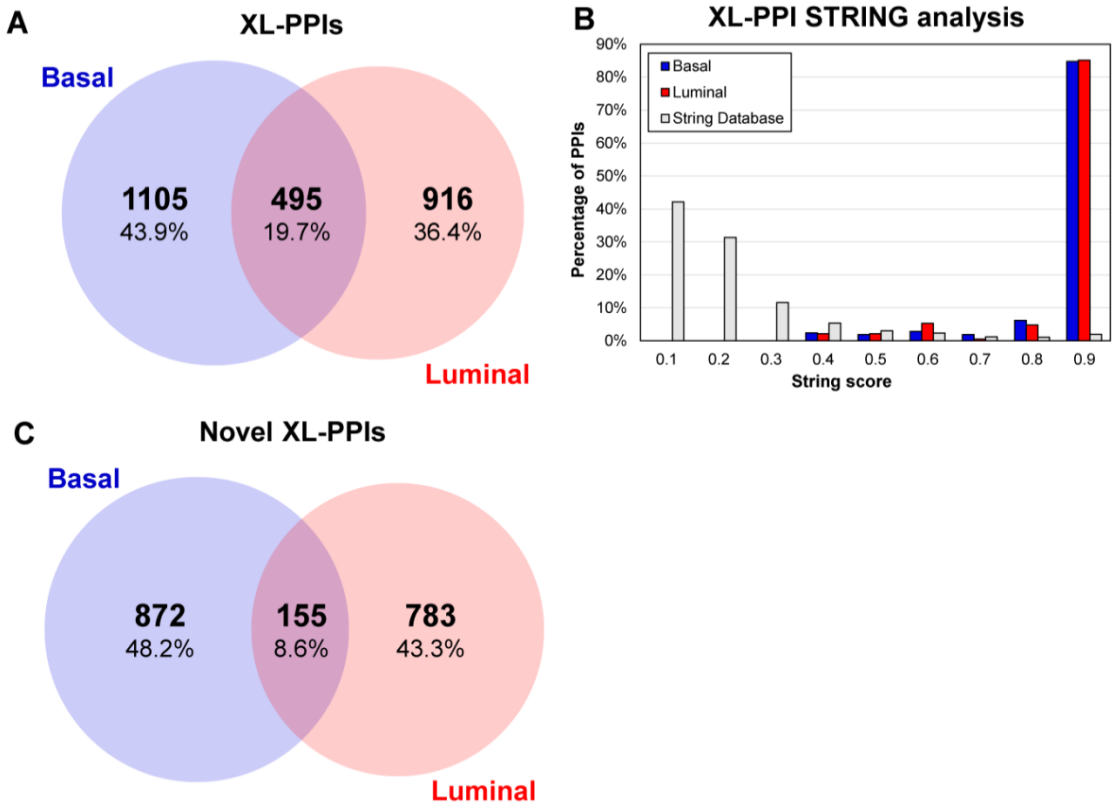

Figure S4. Comparison of PPIs identified from basal and luminal breast cancer PDX samples and their STRING score distribution. (A) Overlap of identified PPIs, **(B)** STRING score distribution of XL-PPIs, and **(C)** overlap of identified novel PPIs.

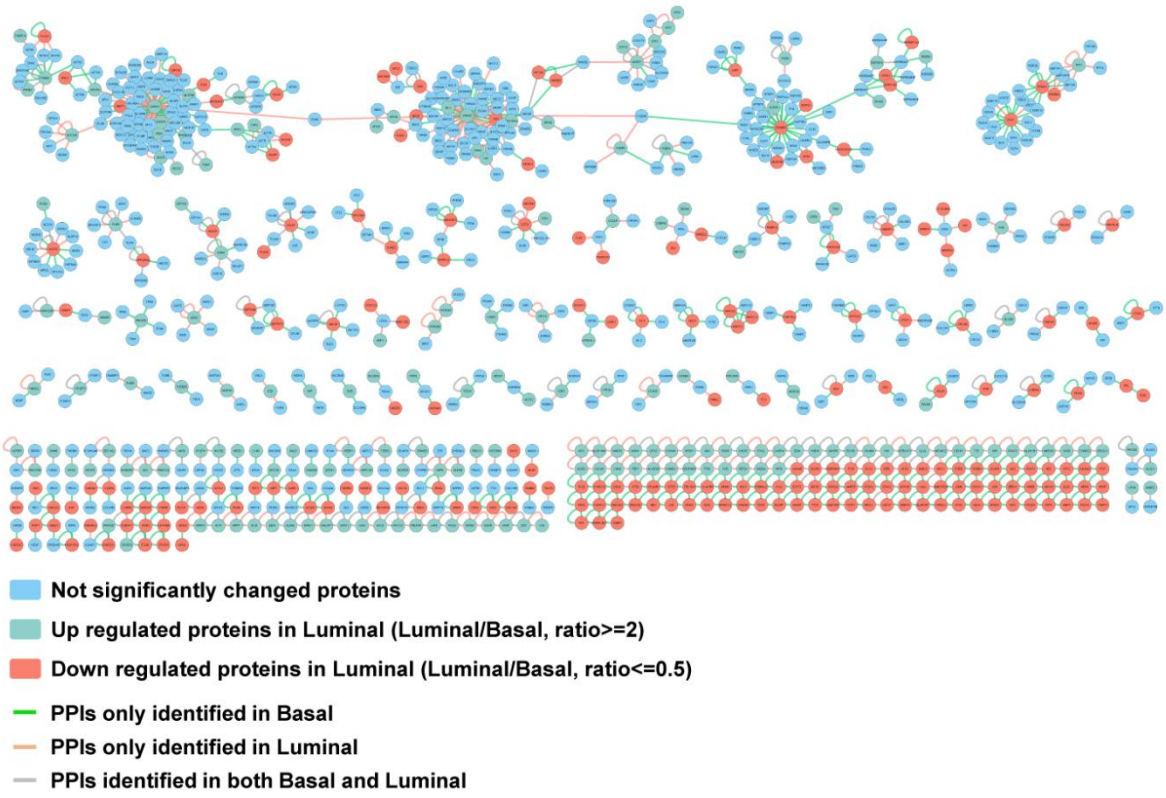

Figure S5. Comparison of XL-PPIs from differential regulated proteins. XL-PPI network of DSBSO cross-linked basal and luminal breast cancer PDX samples. The DSBSO XL-PPI network is composed of 821 nodes and 835 edges. Green edges: PPIs identified in basal breast cancer PDX sample only; red edges: PPIs identified in luminal breast cancer PDX sample only; grey edges: PPIs identified in both basal and luminal breast cancer PDX samples. Blue nodes: proteins not upregulated in either PDX; green nodes: proteins upregulated in luminal breast cancer PDX sample (luminal/basal, ratio>=2); red nodes: proteins downregulated in luminal breast cancer PDX sample (luminal/basal, ratio<=0.5).

References

(1) Wheat, A.; Yu, C.; Wang, X.; Burke, A. M.; Chemmama, I. E.; Kaake, R. M.; Baker, P.; Rychnovsky, S. D.; Yang, J.; Huang, L. Protein interaction landscapes revealed by advanced in vivo cross-linking–mass spectrometry *Proceedings of the National Academy of Sciences* **2021**, 118 (32), e2023360118.

(2) Kao, A.; Chiu, C. L.; Vellucci, D.; Yang, Y.; Patel, V. R.; Guan, S.; Randall, A.; Baldi, P.; Rychnovsky, S. D.; Huang, L. Development of a novel cross-linking strategy for fast and accurate identification of cross-linked peptides of protein complexes *Molecular & Cellular Proteomics* **2011**, 10 (1).

(3) Yang, B.; Wu, Y. J.; Zhu, M.; Fan, S. B.; Lin, J.; Zhang, K.; Li, S.; Chi, H.; Li, Y. X.; Chen, H. F.; Luo, S. K.; Ding, Y. H.; Wang, L. H.; Hao, Z.; Xiu, L. Y.; Chen, S.; Ye, K.; He, S. M.; Dong, M. Q. Identification of cross-linked peptides from complex samples *Nature methods* **2012**, 9 (9): 904-906.

(4) Mertins, P.; Tang, L. C.; Krug, K.; Clark, D. J.; Gritsenko, M. A.; Chen, L.; Clauser, K. R.; Clauss, T. R.; Shah, P.; Gillette, M. A. Reproducible workflow for multiplexed deep-scale proteome and phosphoproteome analysis of tumor tissues by liquid chromatography–mass spectrometry *Nature protocols* **2018**, 13 (7), 1632-1661.

(5) Chen, L.; Zhang, B.; Schnaubelt, M; Shah, P.; Aiyetan, P.; Chan, D.; Zhang, H.; Zhang, Z. MS-PyCloud: An open-source, cloud computing-based pipeline for LC-MS/MS data analysis. *BioRxiv* May 13, **2018**, 320887. DOI: 10.1101/320887.

(6) Fagerberg, L.; Hallström, B. M.; Oksvold, P.; Kampf, C.; Djureinovic, D.; Odeberg, J.; Habuka, M.; Tahmasebpoor, S.; Danielsson, A.; Edlund, K. Analysis of the human tissue-specific expression by genome-wide integration of transcriptomics and antibody-based proteomics. *Molecular & cellular proteomics* **2014**, 13 (2), 397-406.

(7) Tyanova, S.; Albrechtsen, R.; Kronqvist, P.; Cox, J.; Mann, M.; Geiger, T. Proteomic maps of breast cancer subtypes. *Nature communications* **2016**, 7 (1), 10259.

(8) Knowlton, J. J.; Gestaut, D.; Ma, B.; Taylor, G.; Seven, A. B.; Leitner, A.; Wilson, G. J.; Shanker, S.; Yates, N. A.; Prasad, B. V. Structural and functional dissection of reovirus capsid folding and assembly by the prefoldin-TRiC/CCT chaperone network. *Proceedings of the National Academy of Sciences* **2021**, 118 (11), e2018127118.