PhaSepDB: A Database of Liquid-liquid Phase Separation Related Proteins

Kaiqiang You[1,†], Qi Huang[1, †], Chunyu Yu[1, †], Boyan Shen[1], Cristoffer Sevilla[2], Minglei Shi[3], Henning Hermjakob[2,4,*], Yang Chen[3,*], Tingting Li[1,*]

[1] Department of Biomedical Informatics, Peking University Health Science Center, Beijing 100191, China

[2] European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Wellcome Genome Campus, Hinxton, Cambridgeshire CB10 1SD, UK.

[3] MOE Key Laboratory of Bioinformatics, Bioinformatics Division and Center for Synthetic & Systems Biology, BNRist, Department of Automation, Tsinghua University, Beijing 100084, China

[4] State Key Laboratory of Proteomics, Beijing Proteome Research Center, National Center for Protein Sciences (Beijing), Beijing Institute of Life Omics, Beijing 102206, China

* To whom correspondence should be addressed. Tel: +86 13810582006; Email: litt@hsc.pku.edu.cn

Correspondence may also be addressed to Yang Chen. Email: yc@mail.tsinghua.edu.cn

Correspondence may also be addressed to Henning Hermjakob. Email: hhe@ebi.ac.uk

† These authors are equal contributors.

SUPPLEMENTARY DATA

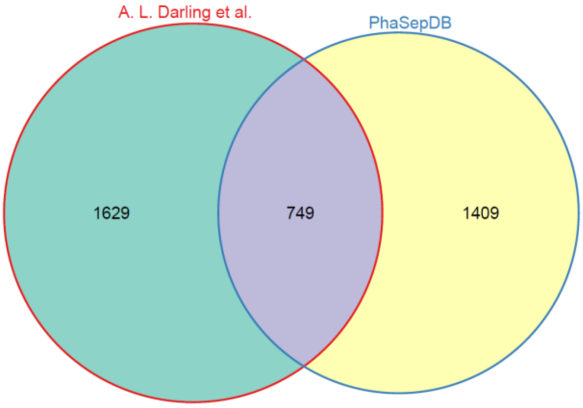

Figure S1 Comparison between PhaSepDB and PMLO-related proteins (A. L. Darling *et al*.). related to Figure 2B

A. L. Darling et al. collect human proteinaceous membrane-less organelles (PMLOs) related proteins from online QuickGO tools complemented with literature search (7). The comparison included proteins localized in nucleolus, nuclear speckles, stress granules, paraspeckles, PML bodies, Cajal body, P granule, P-body, nuclear pore complex and histone locus body, as in Figure 2B.

Table S1 The detail program parameters and annotation resources used in Figure 1B.

The analysis was conducted on Linux operating system with corresponding packages. Note that online server of ESpritz, PLAAC and CIDER should produce identical results.

Type	Program/database	Note
Post-translational modifications (PTMs)	PhosphoSitePlus (1)	Acetylation, methylation, SUMOylation and phosphorylation sites of each proteins were retrieved directly form the PhosphoSitePlus online website (https://www.phosphosite.org/staticDownloads).
Secondary structure and domain annotation	UniProt (2)	The secondary structure annotations were extracted form the UniProt database using the text format . In detail, in the FT (Feature Table) lines, HELIX/STRAND/TURN(secondary structure) , DOMAIN(protein domain), ZN_FING(zinc finger), DNA_BIND(DNA binding), COILED(coiled coil) and COMPBIAS(compositionally biased region) were extracted.
IDR prediction	ESpritz (3) (http://protein.bio.unipd.it/espritz/)	The X-ray prediction type and 5% False Positive Rate (FPR) were used.
Prion like amino acid composition prediction	PLAAC (4)(http://plaac.wi.mit.edu/)	The relative weighting of background probabilities (a) was set as 0.5 and the background frequency was Homo Sapiens(https://github.com/whitehead/plaac/blob/master/web/bg_freqs/bg_freqs_HUMAN.txt).
Pi-pi contact frequency	Pi-Pi (5) (https://doi.org/10.7554/eLife.31486.022)	The default parameters were used in the python scripts given in the original research.
Hydropathy	CIDER (6) (http://pappulab.wustl.edu/CIDER/analysis/)	localCIDER provides an implicit distributed computing model to all of CIDER’s calculations. The hydropathy was calculated for each protein sequence using the function get_linear_hydropathy(blobLen=9) from localCIDER python package.

REFERENCES

1. Hornbeck,P. V, Zhang,B., Murray,B., Kornhauser,J.M., Latham,V. and Skrzypek,E. (2015) PhosphoSitePlus, 2014: mutations, PTMs and recalibrations. *Nucleic Acids Res*, **43**, D512-20.
2. Bateman,A., Martin,M.J., O'Donovan,C., Magrane,M., Alpi,E., Antunes,R., Bely,B., Bingley,M., Bonilla,C., Britto,R., *et al.* (2017) UniProt: the universal protein knowledgebase. *Nucleic Acids Res.*, **45**, D158–D169.
3. Walsh,I., Martin,A.J.M., Di Domenico,T. and Tosatto,S.C.E. (2012) ESpritz: accurate and fast prediction of protein disorder. *Bioinformatics*, **28**, 503–509.
4. Lancaster,A.K., Nutter-Upham,A., Lindquist,S. and King,O.D. (2014) PLAAC: a web and command-line application to identify proteins with prion-like amino acid composition. *Bioinformatics*, **30**, 2501–2502.
5. Vernon,R.M., Chong,P.A., Tsang,B., Kim,T.H., Bah,A., Farber,P., Lin,H. and Forman-Kay,J.D. (2018) Pi-Pi contacts are an overlooked protein feature relevant to phase separation. *Elife*, **7**.
6. Holehouse,A.S., Das,R.K., Ahad,J.N., Richardson,M.O.G. and Pappu,R. V. (2017) CIDER: Resources to Analyze Sequence-Ensemble Relationships of Intrinsically Disordered Proteins. *Biophys. J.*, **112**, 16–21.
7. Darling,A.L., Liu,Y., Oldfield,C.J. and Uversky,V.N. (2018) Intrinsically Disordered Proteome of Human Membrane-Less Organelles. *Proteomics*, **18**, 1–12.