

UCLA

UCLA Electronic Theses and Dissertations

Title

Computational Methods for Studying Protein-Protein Interaction and Association Experiments

Permalink

<https://escholarship.org/uc/item/2h35s75b>

Author

Velasquez, Erick Francisco

Publication Date

2021

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA
Los Angeles

**Computational Methods for Studying Protein-Protein Interaction and Association
Experiments**

A dissertation submitted in partial satisfaction of the requirements for the degree

Doctor of Philosophy

in Biochemistry, Molecular and Structural Biology

by

Erick Velasquez

2021

© Copyright by

Erick Velasquez

2021

ABSTRACT OF THE DISSERTATION

**Computational Methods for Studying Protein-Protein Interactions and
Association Experiments**

by

Erick Velasquez

Doctor of Philosophy in Biochemistry, Molecular and Structural Biology

University of California, Los Angeles, 2021

Professor Jorge Torres, Chair

The elucidation of a protein's interaction/association network is important for defining its biological function. Mass spectrometry-based proteomic approaches have emerged as powerful tools for identifying protein-protein interactions (PPIs) and protein-protein associations (PPAs). However, interactome/association experiments are difficult to interpret considering the complexity and abundance of data that is generated. Although tools have been developed to quantitatively identify protein interactions/associations, there is still a pressing need for easy-to-use tools that allow users to contextualize their results.

To address this, we developed CANVS, a computational pipeline that cleans, analyzes, and visualizes mass spectrometry-based interactome/association data. CANVS is wrapped as an interactive Shiny dashboard, allowing users to easily interface with the pipeline. With simple requirements, users can analyze complex experimental data and create PPI/A networks. The application integrates systems biology databases like BioGRID and CORUM to contextualize the results. Furthermore, CANVS features a Gene Ontology tool that allows users to identify relevant GO terms in their results and create visual networks with proteins associated with relevant GO terms. As examples, we recently used the analytical framework included in CANVS to study the PPI/A networks of DUSP7, which helped to define its regulation of ERK2 during mitosis and also to analyze the PPA networks of core spindle assembly checkpoint proteins. Overall, CANVS is an easy-to-use application that benefits all researchers, especially those who lack an established bioinformatic pipeline and are interested in studying interactome/association data.

Additionally, we describe a supervised machine learning method that incorporates annotated data from the contaminant repository for affinity purification data (CRAPome) that predicts contaminants in affinity and proximity purification data. The method involves first calculating amino acid content, sequence order, hydrophobicity and hydrophilicity from protein sequence. Then balancing data using data augmentation methods. Finally, measuring precision and accuracy using protein-protein interaction/association data. The results suggest that our supervised method can predict with 90% accuracy contaminants in protein-protein interaction/association data.

The dissertation of Erick Velasquez is approved.

James Akira Wohlschlegel

Joseph Ambrose Loo

Catherine F. Clarke

Jorge Torres, Committee Chair

University of California, Los Angeles

2021

DEDICATION

I would like to dedicate this work to my family, especially my mother Angelica Velasquez who is my biggest inspiration.

Me gustaría dedicar este trabajo a mi familia, especialmente a mi madre Angélica Velásquez, quien es mi mayor inspiración.

CONTENTS

<u>CHAPTER</u>	<u>PAGE</u>
ABSTRACT OF DISSERTATION.....	ii-iii
COMMITTEE PAGE.....	iv
DEDICATION.....	v
CONTENTS.....	vi
ACKNOWLEDGEMENTS.....	vii
CURRICULUM VITAE.....	viii-ix
CHAPTER 1 – Background.....	1
References.....	7
CHAPTER 2 – CANVS; an easy-to-use application for the analysis and visualization of MS-based PPI/A data.....	13
References.....	35
CHAPTER 3 – Mapping Proximity Associations of Spindle Assembly Checkpoint Proteins.....	42
References.....	81
CHAPTER 4 – Future Direction: Predicting Contaminants in PPI/A Data Using Supervised Machine Learning.....	92
References.....	106
Appendix Chapter 1 – CANVS Instruction Manual.....	110
Appendix Chapter 2 – Mapping Proximity Associations of Spindle Assembly Checkpoint Proteins Supplemental.....	123
Appendix Chapter 3 – DUSP7 Regulates the Activity of ERK2 to Promote Proper Chromosome Alignment During Cell.....	143
Appendix Chapter 4 – The myosin regulatory light chain Myl5 localizes to mitotic spindle poles and is required for proper cell division.....	152
Appendix Chapter 5 – Regulation of Iron Homeostasis through Parkin-Mediated Lactoferrin Ubiquitylation.....	165
Appendix Chapter 6 – Leukemia Cell Cycle Chemical Profiling Identifies the G2-phase Leukemia Specific Inhibitor Leusin-1.....	171

ACKNOWLEDGEMENTS

I would like to acknowledge the amazing individuals who have made the work described in this thesis possible. First, I would like to thank my amazing mentor and supervisor Dr. Jorge Torres for his incredible support and guidance. He is truly an exceptional mentor, and someone that I am honored to have in my life. He has truly changed my life outcomes and for that I will be forever thankful. Second, I would like to thank the Torres lab members, specially Dr. Yenni Garcia and Ankur Gholkar. Not only was it a pleasure working with them every day but they also took the time to mentor me and provide counsel when I needed it the most. Third, I would like to thank my committee members; Dr. James Wohlschlegel, Dr. Cathy Clarke, and Dr. Joseph Loo. Throughout my years in graduate school they have been able to provide advice and challenged me to grow a scientist and an individual. Lastly, I would like to thank the undergraduates and graduate students I have mentored April Morales, Nicole Lynn, and Kayla Rich.

CURRICULUM VITAE

Erick Francisco Velasquez

EDUCATION

University of California Los Angeles, Los Angeles CA. September 2016 – Present

Ph.D. Candidate, Biochemistry, Molecular and Structural Biology GPA: 3.9/4.0

Pomona College, Claremont CA. September 2012 – May 2016

B.A., Chemistry, GPA: 3.5/4.0

Cambridge University, United Kingdom January 2015 – June 2015

Study Abroad

SKILLS

R, Shiny-App Development, Python, MySQL query, Jupyter Notebooks, Significance Statistics, Machine Learning, Plotting with ggplot2, Human Cell Culture, Molecular Cloning, Mass Spectrometry, Molecular Docking

RESEARCH EXPERIENCE

Ph.D. Candidate, Biochemistry, Molecular and Structural Biology June 2017– Present

- Applied significance statistics and machine learning algorithms to mass spectrometry data (labeled and label-free), while data scrapping large biological databases to contextualize novel protein-protein and protein-drug interactions.

Bioinformatics Intern, Genentech, Dept. of Biomarker Development July 2019 – January 2020

- Developed computational pipeline to automate quality control processes on targeted proteomic experiments using unsupervised machine learning methods.

Protein Chemistry Intern, Genentech, Dept. of Protein Chemistry May 2016– August 2016

- Used rational design, modeled and synthesized cleavable linkers, and tested conjugates in human cell cultures.

Research Assistant, Pomona College, Dept. of Chemistry (Liu Lab) March 2014 – May 2016

- Used classical molecular biology techniques to develop large RNA libraries. Optimized cell sorting to select for populations that expressed riboswitches of interest.

Research Assistant, Tufts University, Dept. of Chemistry (Walt Lab) May 2013 – August 2013

- Optimized single molecule ELISA to quantify biomarkers at femtomolar concentrations.

Research Assistant, Pomona College, Dept. of Chemistry (Garza Lab) November 2012 – May 2013

- Incorporated molecular dynamics algorithms to study molecular diffusion patterns in different fractals.

AWARDS AND RECOGNITION

Howard Hughes Medical Institute Gilliam Fellowship	(June 2019 – National)
Genentech Leader Intern Exchange	(May 2019 – Corporate)
Hanson-Dow Excellence in Teaching Award	(April 2019 – Department)
UCLA Molecular Biology Whitcome Fellowship	(June 2018 – School)
Ford Foundation Fellowship Alternate and Honorable Mention	(May 2017, 2018 – National)
UCLA Chemistry Biology Interface Associate	(May 2017 – School)
Eugene V. Cota Robles Fellowship	(May 2016 – Department)
American Chemical Society’s Organic Chemistry Undergraduate Award	(February 2016 – National)
Annual Biomedical Research Conference for Minority Students Award	(November 2015 – National)
QuestBridge National College Match Recipient	(December 2011– National)

PUBLICATIONS

Velasquez E.F., Lee J., Matthews R., Eshghi S. “Automated Quality Assessment of Chromatographic Peaks in Targeted Proteomics Data using Unsupervised Machine Learning”. In preparation

Velasquez E.F., Garcia Y.A, Ramirez I., Gholkar A.A., Torres J.Z. “CANVS: an easy-to-use application for the analysis and visualization of mass spectrometry-based protein-protein interaction/association data.” *Molecular biology of the cell*, mbcE21050257. 25 Aug. 2021, doi:10.1091/mbc.E21-05-0257.

Garcia Y.A , **Velasquez E.F.**, Ramirez I., Filbert N.G., Williams-Hamilton T., Perez L., Gholkar A.A., Cohn W., Gao L.W., Capri J., Whitelegge J.P., and Torres J.Z. “Mapping Proximity Associations of Core Spindle Assembly Checkpoint Proteins.” *Journal of proteome research* vol. 20,7 (2021): 3414-3427. doi:10.1021/acs.jproteome.0c00941.

Guo X., Garcia Y.A., Ramirez I., **Velasquez E.F.**, Gao L.W., Gholkar A.A., Whitelegge J., Tofig B., Damoiseaux R., and Torres, J.Z. “DUSP7 Regulates the Activity of ERK2 to Promote Proper Chromosome Alignment During Cell Division”. *Journal of Biological Chemistry* 2021, Apr 15:100676. doi: 10.1016/j.jbc.2021.100676.

Ramirez I., Gholkar A.A., **Velasquez E.F.**, Guo X., Tofig B., Damoiseaux R., Torres J.Z. The myosin regulatory light chain Myl5 localizes to mitotic spindle poles and is required for proper cell division. *Cytoskeleton (Hoboken)*. 2021 Feb;78(2):23-35. doi: 10.1002/cm.21654.

Gholkar A.A., Schmollinger S., **Velasquez E.F.**, Lo Y.C., Cohn W., Capri J., Dharmarajan H., Deardorff W.J., Gao L.W., Abdusamad M., Whitelegge J.P., Torres J.Z.. “Regulation of Iron Homeostasis through Parkin-Mediated Lactoferrin Ubiquitylation. *Biochemistry*”. 2020 Aug 18;59(32):2916-2921. doi: 10.1021/acs.biochem.0c00504.

Scales S., Tsai S. P., Zacharias N., Cruz-Chuh J., Bullen G., **Velasquez E.F.**, Chang J., Bruguera E., Kozak K. R., Sadowsky J., “Development of a Cysteine-Conjugatable Disulfide FRET Probe: Influence of Charge on Linker Cleavage and Payload Trafficking for an Anti-HER2 Antibody Conjugate.” *Bioconjugate Chem.* 30, 3046–3056 (2019).

Xia X., Lo Y.C., Gholkar A.A., Senese S., Ong J., **Velasquez E.F.**, Damoiseaux R., and Torres J.Z. “Leukemia Cell Cycle Chemical Profiling Identifies the G2-phase Leukemia Specific Inhibitor Leusin-1”. *ACS Chem. Biol.* 14, 994–1001 (2019).

CHAPTER 1

Background

1.1 Protein-protein interactions (PPIs) and indirect protein-protein associations (PPAs)

Direct protein-protein interactions (PPIs) and indirect protein-protein associations (PPAs) are critical to understanding the biological function of a protein of interest (POI). PPI/As can inform on how a POI is compartmentalized within the cell, how it forms higher-order complexes, how it is regulated, and how it coordinates with other proteins to carry out specific cellular processes in a spatial and temporal manner (Yugandhar *et al.*, 2019; Lu *et al.*, 2020). More broadly PPI/A networks have been used to analyze the composition of cellular structures like centrosomes, kinetochores, cilia, and other organelles, and to define the function of cell cycle proteins (Torres *et al.*, 2011; Firat-Karalar and Stearns, 2015; Cheung *et al.*, 2016; Go *et al.*, 2019; Remnant *et al.*, 2019; Garcia *et al.*, 2021). Most importantly, PPI/As play important roles in driving key cellular events like DNA replication, transcription, translation and signal transduction (Koh *et al.*, 2012; Lu *et al.*, 2020).

Hence, thoroughly understanding how PPI/As give rise to cellular events allows scientists to define how aberrant PPI/As can give rise to disease (White *et al.*, 2008; Blazer and Neubig, 2009) and how they can be used to develop novel therapeutics (Rosell and Fernández-Recio, 2018). Classic drug targets involve enzymes, ion channels or receptors making PPI/As an important modality that can be used as therapeutic targets (Milroy *et al.*, 2014). One of the main difficulties in small molecule discovery is large contact surfaces involved in PPI/As compared to protein-small molecule interactions, therefore creating therapeutic modalities that mimic PPI/As has become an attractive option in

therapeutic development (Hill *et al.*, 2014; Nevola and Giralt, 2015). PPI/As as intervention targets have been proven to be useful in generating novel treatments for refractory diseases demonstrating promising capabilities as novel therapeutic targets (Nero *et al.*, 2014).

1.2 Approaches to identify PPI/As

Studying PPI/As is important to developing a better understanding of biological systems and to developing novel therapeutics, however first scientists have to identify probable interactions and associations. Several approaches have been used to identify PPIs including yeast two-hybrid (Fields and Song, 1989; Chien *et al.*, 1991), fluorescent resonance energy transfer (Selvin, 1995), and protein fragment complementation assay (Michnick *et al.*, 2000). Although these classical methods are important for identifying and validating PPIs, mass spectrometry-based approaches have made high-throughput identification of PPIs possible (Yugandhar *et al.*, 2019). Affinity purification mass spectrometry (AP-MS) has become the conventional method to identify PPIs, since it isolates protein complexes from cell lysates in near physiological conditions (Gingras *et al.*, 2007). More recently the field has transitioned to defining PPAs, which may represent direct protein-protein interactions or local protein neighborhoods, through proximity-dependent biotinylation methods (Perkins *et al.*, 2010). A popular method is BioID, where a protein of interest (POI) is tagged with a promiscuous biotin ligase, which biotinylates proteins in close proximity to the POI in the presence of biotin (Sears *et al.*, 2019).

Exploratory PPI/A Experimental Pipeline

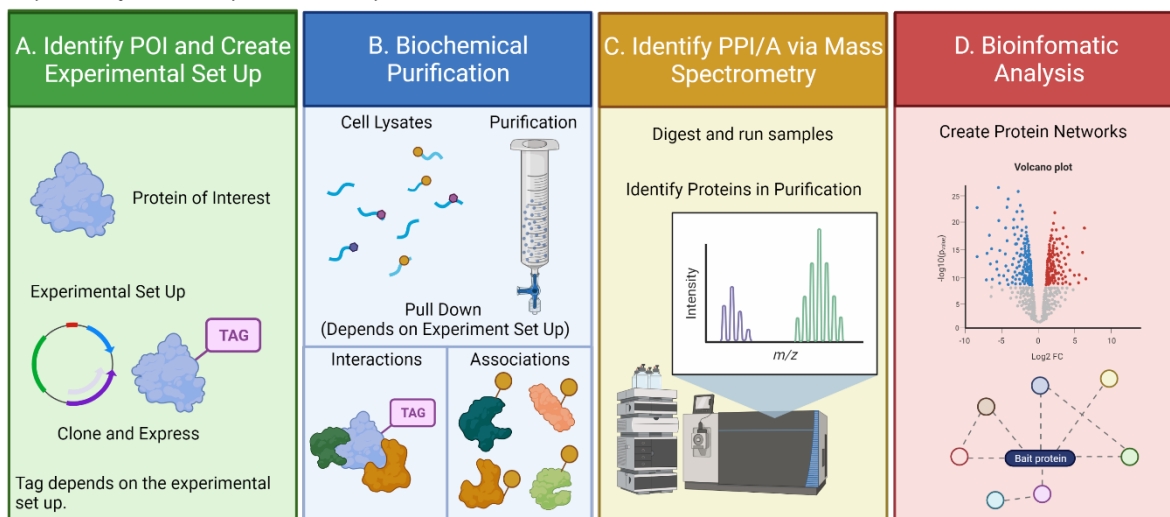


Figure 1. Exploratory PPI/A Experimental Pipeline. A. Identify protein of interest and create an experimental set up that seeks to identify either protein-protein interactions or protein-protein associations. B. Biochemical purification. Based on type of experimental set up, pull POI or proteins that have been tagged through close proximity of POI. C. Identify proteins present in pull down purification via mass spectrometry. D. Bioinformatic analysis of proteins identified via mass spectrometry. If quantitative protein values are available with multiple replicates significant change can be calculated and PPI/As can be represented in protein networks. Created with BioRender.com

1.3 Analysis workflow of MS-based PPI/A experiments

Overall, popular pipelines for PPI/A approaches involve identifying a POI, tagging the POI with an appropriate protein tag (Figure 1A), expression of the tagged-POI, biochemical purifications (Figure 1B), MS of purifications to identify proteins (Figure 1C), and qualitative/quantitative bioinformatic analyses of identified proteins (Figure 1D) (Sears *et*

al., 2019; Yugandhar *et al.*, 2019). Mass spectrometry-based approaches to identify PPI/As generate large amounts of PPI/A data, which are difficult to analyze and interpret. To overcome these issues, computational tools have been developed to quantitatively analyze PPI/A data (Choi *et al.*, 2011; Choi *et al.*, 2012; Nesvizhskii, 2012; Teo *et al.*, 2014), to create visual representations of the analyses (Knight *et al.*, 2017), and to generate PPI/A networks (Shannon *et al.*, 2003; Szklarczyk *et al.*, 2019). One pipeline, APOSTL, integrates these steps and automates the process within a Galaxy framework (Kuenzi *et al.*, 2016). However, currently APOSTL does not allow users to filter search results by Gene Ontology-based (GO) terms or integrate protein complex data. Furthermore, most PPI/A data analysis computational tools focus on the accuracy of protein identification, instead of how the identified proteins might be associated at the molecular level.

1.4 PPI/A Network Biology Databases

PPI/A MS-based experiments seek to unravel network data that allows scientists to understand how proteins might contribute to complex biological events. To better aid scientists and contextualize their results many large biological datasets have been arranged that describe protein interactions, associations, and function. Most importantly, these databases are set up to help scientists derive hypotheses about cellular function in the context of a protein or proteins of interest. Of particular interest are databases that group proteins as part of a complex, detail the function or location of proteins and show previously known experimental information about protein interactions/associations. The comprehensive resource of mammalian protein complexes (CORUM) is a manually

curated repository for experimentally characterized protein complexes in mammalian organisms, providing the necessary information to identify common complexes in PPI/A experimental data (Giurgiu *et al.*, 2019). The gene ontology resource (GO Consortium) is the largest source of information about gene and gene product attributes across species. GO data allows scientists to study PPI/A result within the context of their molecular function, cellular component or whether they aid in a certain biological process (Ashburner *et al.*, 2000; Consortium, 2019). The biological general repository of interaction datasets (BioGRID) is a database that archives genetic and protein interaction information (Oughtred *et al.*, 2019). BioGRID allows scientists to contextualize their results based on previously identified interactions so that scientist gain more confidence about possible interactions and associations. These three databases can be used individually or together to create a more thorough understanding of PPI/A experiments, which is particularly helpful in the exploratory stages of the pipeline.

REFERNCES

1. Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., Harris, M.A., Hill, D.P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J.C., Richardson, J.E., Ringwald, M., Rubin, G.M., and Sherlock, G. (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nature genetics* 25.
2. Blazer, L.L., and Neubig, R.R. (2009). Small molecule protein-protein interaction inhibitors as CNS therapeutic agents: current progress and future hurdles. *Neuropsychopharmacology : official publication of the American College of Neuropsychopharmacology* 34.
3. Cheung, K., Senese, S., Kuang, J., Bui, N., Ongpipattanakul, C., Gholkar, A., Cohn, W., Capri, J., Whitelegge, J.P., and Torres, J.Z. (2016). Proteomic Analysis of the Mammalian Katanin Family of Microtubule-severing Enzymes Defines Katanin p80 subunit B-like 1 (KATNBL1) as a Regulator of Mammalian Katanin Microtubule-severing. *Molecular & cellular proteomics : MCP* 15.
4. Chien, C., Bartel, P., Sternglanz, R., and Fields, S. (1991). The two-hybrid system: a method to identify and clone genes for proteins that interact with a protein of interest. *Proceedings of the National Academy of Sciences of the United States of America* 88.
5. Choi, H., Larsen, B., Lin, Z.-Y., Breitkreutz, A., Mellacheruvu, D., Fermin, D., Qin, Z.S., Tyers, M., Gingras, A.-C., and Nesvizhskii, A.I. (2011). SAINT: probabilistic scoring of affinity purification-mass spectrometry data. *Nature methods* 8.

6. Choi, H., Liu, G., Mellacheruvu, D., Tyers, M., Gingras, A.-C., and Nesvizhskii, A.I. (2012). Analyzing protein-protein interactions from affinity purification-mass spectrometry data with SAINT. *Current protocols in bioinformatics Chapter 8*.
7. Consortium, T.G.O. (2019). The Gene Ontology Resource: 20 years and still GOing strong. *Nucleic acids research 47*.
8. Fields, S., and Song, O. (1989). A novel genetic system to detect protein-protein interactions. *Nature 340*.
9. Firat-Karalar, E.N., and Stearns, T. (2015). Probing mammalian centrosome structure using BiLD proximity-dependent biotinylation. *Methods in cell biology 129*.
10. Garcia, Y.A., Velasquez, E.F., Gao, L.W., Gholkar, A.A., Clutario, K.M., Cheung, K., Williams-Hamilton, T., Whitelegge, J.P., and Torres, J.Z. (2021). Mapping Proximity Associations of Core Spindle Assembly Checkpoint Proteins. *Journal of proteome research 20*.
11. Gingras, A., Gstaiger, M., Raught, B., and Aebersold, R. (2007). Analysis of protein complexes using mass spectrometry. *Nature reviews. Molecular cell biology 8*.
12. Giurgiu, M., Reinhard, J., Brauner, B., Dunger-Kaltenbach, I., Fobo, G., Frishman, G., Montrone, C., and Ruepp, A. (2019). CORUM: the comprehensive resource of mammalian protein complexes-2019. *Nucleic acids research 47*.
13. Go, C.D., Knight, J.D.R., Rajasekharan, A., Rathod, B., Hesketh, G.G., Abe, K.T., Youn, J.-Y., Samavarchi-Tehrani, P., Zhang, H., Zhu, L.Y., Popiel, E., Lambert, J.-P., Coyaud, É., Cheung, S.W.T., Rajendran, D., Wong, C.J.,

- Antonicka, H., Pelletier, L., Raught, B., Palazzo, A.F., Shoubridge, E.A., and Gingras, A.-C. (2019). A proximity biotinylation map of a human cell.
14. Hill, T.A., Shepherd, N.E., Diness, F., and Fairlie, D.P. (2014). Constraining cyclic peptides to mimic protein structure motifs. *Angewandte Chemie (International ed. in English)* 53.
15. Knight, J.D.R., Choi, H., Gupta, G.D., Pelletier, L., Raught, B., Nesvizhskii, A.I., and Gingras, A.-C. (2017). ProHits-viz: a suite of web tools for visualizing interaction proteomics data. *Nature methods* 14.
16. Koh, G.C.K.W., Porras, P., Aranda, B., Hermjakob, H., and Orchard, S.E. (2012). Analyzing protein-protein interaction networks. *Journal of proteome research* 11.
17. Kuenzi, B.M., Borne, A.L., Li, J., Haura, E.B., Eschrich, S.A., Koomen, J.M., Rix, U., and Stewart, P.A. (2016). APOSTL: An Interactive Galaxy Pipeline for Reproducible Analysis of Affinity Proteomics Data. *Journal of proteome research* 15.
18. Lu, H., Zhou, Q., He, J., Jiang, Z., Peng, C., Tong, R., and Shi, J. (2020). Recent advances in the development of protein-protein interactions modulators: mechanisms and clinical trials. *Signal transduction and targeted therapy* 5.
19. Michnick, S., Remy, I., Campbell-Valois, F., Vallée-Bélisle, A., and Pelletier, J. (2000). Detection of protein-protein interactions by protein fragment complementation strategies. *Methods in enzymology* 328.
20. Milroy, L.-G., Grossmann, T.N., Hennig, S., Brunsveld, L., and Ottmann, C. (2014). Modulators of protein-protein interactions. *Chemical reviews* 114.

21. Nero, T.L., Morton, C.J., Holien, J.K., Wielens, J., and Parker, M.W. (2014). Oncogenic protein interfaces: small molecules, big challenges. *Nature reviews. Cancer* 14.
22. Nesvizhskii, A.I. (2012). Computational and informatics strategies for identification of specific protein interaction partners in affinity purification mass spectrometry experiments. *Proteomics* 12.
23. Nevola, L., and Giralt, E. (2015). Modulating protein-protein interactions: the potential of peptides. *Chemical communications (Cambridge, England)* 51.
24. Oughtred, R., Stark, C., Breitkreutz, B.-J., Rust, J., Boucher, L., Chang, C., Kolas, N., O'Donnell, L., Leung, G., McAdam, R., Zhang, F., Dolma, S., Willems, A., Coulombe-Huntington, J., Chatr-Aryamontri, A., Dolinski, K., and Tyers, M. (2019). The BioGRID interaction database: 2019 update. *Nucleic acids research* 47.
25. Perkins, J.R., Diboun, I., Dessailly, B.H., Lees, J.G., and Orengo, C. (2010). Transient protein-protein interactions: structural, functional, and network properties. *Structure (London, England : 1993)* 18.
26. Remnant, L., Booth, D.G., Vargiu, G., Spanos, C., Kerr, A.R.W., and Earnshaw, W.C. (2019). In vitro BioID: mapping the CENP-A microenvironment with high temporal and spatial resolution. *Molecular biology of the cell* 30.
27. Rosell, M., and Fernández-Recio, J. (2018). Hot-spot analysis for drug discovery targeting protein-protein interactions. *Expert opinion on drug discovery* 13.

28. Sears, R.M., May, D.G., and Roux, K.J. (2019). BioID as a Tool for Protein-Proximity Labeling in Living Cells. *Methods in molecular biology* (Clifton, N.J.) *2012*.
29. Selvin, P. (1995). Fluorescence resonance energy transfer. *Methods in enzymology* *246*.
30. Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome research* *13*.
31. Szklarczyk, D., Gable, A.L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., Simonovic, M., Doncheva, N.T., Morris, J.H., Bork, P., Jensen, L.J., and Mering, C.v. (2019). STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic acids research* *47*.
32. Teo, G., Liu, G., Zhang, J., Nesvizhskii, A.I., Gingras, A.-C., and Choi, H. (2014). SAINTexpress: improvements and additional features in Significance Analysis of INTERactome software. *Journal of proteomics* *100*.
33. Torres, J.Z., Summers, M.K., Peterson, D., Brauer, M.J., James Lee, S.S., Gholkar, A.A., Lo, Y.-C., Lei, X., Jung, K., Anderson, D.C., Davis, D.P., Belmont, L., and Jackson, P.K. (2011). The STARD9/Kif16a kinesin associates with mitotic microtubules and regulates spindle pole assembly. *Cell* *147*.
34. White, A.W., Westwell, A.D., and Brahehi, G. (2008). Protein-protein interactions as targets for small-molecule therapeutics in cancer. *Expert reviews in molecular medicine* *10*.

35. Yugandhar, K., Gupta, S., and Yu, H. (2019). Inferring Protein-Protein Interaction Networks From Mass Spectrometry-Based Proteomic Approaches: A Mini-Review. *Computational and structural biotechnology journal* 17.

CHAPTER 2

CANVS; an easy-to-use application for the analysis and visualization of MS-based PPI/A data

CANVS: an easy-to-use application for the analysis and visualization of mass spectrometry-based protein-protein interaction/association data

Erick F. Velasquez¹, Yenni A. Garcia¹, Ivan Ramirez¹, Ankur A. Gholkar¹ and Jorge Z. Torres^{1-3*}

¹Department of Chemistry and Biochemistry, University of California, Los Angeles, CA 90095, USA

²Molecular Biology Institute, University of California, Los Angeles, CA 90095, USA

³Jonsson Comprehensive Cancer Center, University of California, Los Angeles, CA 90095, USA

***Corresponding author:** Jorge Z. Torres

Email: torres@chem.ucla.edu

Reprinted (adapted) with permission from { Velasquez, Erick F et al. "CANVS: an easy-to-use application for the analysis and visualization of mass spectrometry-based protein-protein interaction/association data." Molecular biology of the cell vol. 32,21 (2021): br9. doi:10.1091/mbc.E21-05-0257}. Copyright {2021} Molecular Biology of the Cell

Running Head: CANVS: MS data analysis and visualization app

Keywords: protein-protein interactions, protein-protein associations, protein-complex annotations, protein network, mass spectrometry

Abbreviations: PPI, protein-protein interaction; PPA, protein-protein association; CANVS, Clean Analyze Network Visualization Software; BioGRID, Biological General Repository for Interaction Datasets; CORUM, Comprehensive Resource of Mammalian Protein Complexes; GO, Gene Ontology; app, application; MS, mass spectrometry

ABSTRACT

The elucidation of a protein's interaction/association network is important for defining its biological function. Mass spectrometry-based proteomic approaches have emerged as powerful tools for identifying protein-protein interactions (PPIs) and protein-protein associations (PPAs). However, interactome/association experiments are difficult to interpret considering the complexity and abundance of data that is generated. Although tools have been developed to quantitatively identify protein interactions/associations, there is still a pressing need for easy-to-use tools that allow users to contextualize their results. To address this, we developed CANVS, a computational pipeline that cleans, analyzes, and visualizes mass spectrometry-based interactome/association data. CANVS is wrapped as an interactive Shiny dashboard, allowing users to easily interface with the pipeline. With simple requirements, users can analyze complex experimental data and create PPI/A networks. The application integrates systems biology databases like BioGRID and CORUM to contextualize the results. Furthermore, CANVS features a Gene Ontology tool that allows users to identify relevant GO terms in their results and create visual networks with proteins associated with relevant GO terms. Overall, CANVS is an easy-to-use application that benefits all researchers, especially those who lack an established bioinformatic pipeline and are interested in studying interactome/association data.

INTRODUCTION

Direct protein-protein interactions (PPIs) and indirect protein-protein associations (PPAs) are critical to understanding the biological function of a protein of interest (POI). PPI/As can inform on how a POI is compartmentalized within the cell, how it forms higher-order complexes, how it is regulated, and how it coordinates with other proteins to carry out specific cellular processes in a spatial and temporal manner (Yugandhar *et al.*, 2019; Lu *et al.*, 2020). More broadly PPI/A networks have been used to analyze the composition of cellular structures like centrosomes, kinetochores, cilia, and other organelles, and to define the function of cell cycle proteins (Torres *et al.*, 2011; Firat-Karalar and Stearns, 2015; Cheung *et al.*, 2016; Go *et al.*, 2019; Remnant *et al.*, 2019; Garcia *et al.*, 2021; Guo *et al.*, 2021). Several approaches have been used to identify PPIs including yeast two-hybrid (Fields and Song, 1989; Chien *et al.*, 1991), fluorescent resonance energy transfer (Selvin, 1995), and protein fragment complementation assay (Michnick *et al.*, 2000). Although these classical methods are important for identifying and validating PPIs, mass spectrometry-based approaches have made high-throughput identification of PPIs possible (Yugandhar *et al.*, 2019). Affinity purification mass spectrometry (AP-MS) has become the conventional method to identify PPIs, since it isolates protein complexes from cell lysates in near physiological conditions (Gingras *et al.*, 2007). More recently the field has transitioned to defining PPAs, which may represent direct protein-protein interactions or local protein neighborhoods, through proximity-dependent biotinylation methods (Perkins *et al.*, 2010). A popular method is BioID, where a protein of interest (POI) is tagged with a promiscuous biotin ligase, which biotinylates proteins in close proximity to the POI in the presence of biotin (Sears *et al.*, 2019). Overall, popular pipelines for PPI/A

approaches involve identifying a POI, tagging the POI with an appropriate protein tag, expression of the tagged-POI, biochemical purifications, MS of purifications to identify proteins, and qualitative/quantitative bioinformatic analyses of identified proteins (Sears *et al.*, 2019; Yugandhar *et al.*, 2019).

Mass spectrometry-based approaches to identify PPI/As generate large amounts of PPI/A data, which are difficult to analyze and interpret. To overcome these issues, computational tools have been developed to quantitatively analyze PPI/A data (Choi *et al.*, 2011; Choi *et al.*, 2012; Nesvizhskii, 2012; Teo *et al.*, 2014), to create visual representations of the analyses (Knight *et al.*, 2017), and to generate PPI/A networks (Shannon *et al.*, 2003; Szklarczyk *et al.*, 2019). One pipeline, APOSTL, integrates these steps and automates the process within a Galaxy framework (Kuenzi *et al.*, 2016). However, currently APOSTL does not allow users to filter search results by Gene Ontology-based (GO) terms or integrate protein complex data. Furthermore, most PPI/A data analysis computational tools focus on the accuracy of protein identification, instead of how the identified proteins might be associated at the molecular level. With this in mind, we sought to develop a computational pipeline that allowed users with no computational background to explore PPI/A data interactively within the context of relevant biological processes, molecular functions, cellular components, and protein-protein-complex interactions.

Here, we present CANVS (Clean Analyze Network Visualization Software), an open access computational pipeline that cleans mass-spectrometry PPI/A data through statistical analyses and annotates identified proteins with proteoinformatic databases like BioGRID (Biological General Repository for Interaction Datasets) (Oughtred *et al.*, 2019)

and CORUM (Comprehensive Resource of Mammalian Protein Complexes) (Giurgiu *et al.*, 2019) to create protein interaction/association networks. Furthermore, CANVS allows users to apply GO (Gene Ontology) (Ashburner *et al.*, 2000; Consortium, 2021) filters to create protein networks relevant to biological processes, cellular locations, or molecular functions of interest. To ensure accessibility to all researchers, CANVS is wrapped in a Golem framework (Fay *et al.*, 2021) and is deployed as a Shiny dashboard app that can be downloaded and installed locally on a Windows system (<https://sourceforge.net/projects/canvs/files/>). CANVS can be used as a stand-alone tool; however, the user can also upload results from other proteomic pipelines to quickly generate protein networks and identify proteins with relevant biological associations to a POI. Overall, CANVS provides an easy-to-use interactive framework where proteoinformatic resources are integrated to quantify, contextualize, and visualize data from PPI/A experiments that enables users to better understand the biological role of their POI.

RESULTS

Features

CANVS is an open-access easy-to-use pipeline for studying protein-protein interaction/association data. CANVS was created so that researchers with no computational background can quickly analyze mass spectrometry data from affinity-based and proximity-based protein purifications, with an emphasis on identifying biologically interesting PPI/As that can be further validated and explored. Briefly, the CANVS pipeline (Figure 1) can be divided into four steps, 1) uploading data, 2) cleaning data to identify significant protein hits, 3) analyzing results by applying proteoinformatic

databases, and 4) visualizing the resulting PPI/A networks. The CANVS pipeline takes advantage of user interface web development packages in R and is wrapped as a Shiny dashboard app that can be installed locally on a Windows system.

1. Data upload/pre-processing

CANVS accepts csv or tab-delimited text files with five columns containing information on: protein UniProt (Apweiler *et al.*, 2004) accession number, score or quantitative value, protein description, file name, and protein bait name (Supplemental Material - Uploading Data). Files are then merged and the merged data table can be accessed in the *Upload* tab (Supplemental Figure S1). Additionally, the user has the option of uploading one data table, for the control and experimental each, that has all the information about replicates/conditions. The data tables are interactive and the user can search for specific keywords in the search section. Pre-processing involves both determining how many replicates a protein needs to appear in to carry out the analysis and normalization across replicates and baits (Figure 2 and Supplemental Figures S2-S4). High-throughput MS-based approaches to identify PPI/As contain systematic biases due to steps in data processing and generation (Chawade *et al.*, 2014). To overcome these biases the field has adopted normalization methods that aim to make samples more comparable across replicates/conditions (Chawade *et al.*, 2014; Välikangas *et al.*, 2018). CANVS gives the user the option to normalize by the median and scales samples so that each purification has the same median value (Välikangas *et al.*, 2018). We recommend using this method if no previous normalization was performed and if the user suspects variation across purifications due to human error (sample preparation) or analytical errors (device

calibration, temperature fluctuations, etc.). Both settings are pre-defined by the user in the *Clean* tab where the user can filter by the number of replicates a protein must be identified in to be considered for further analysis and whether to normalize by the median (Figure 2). Please see Supplemental Material for details and step-by-step user instructions.

2. Semi quantitative/qualitative analysis of protein hits

After setting up the pre-processing options, users can determine whether to perform a statistical analysis to identify significant protein hits (Supplemental Material – Cleaning Data). CANVS uses spectral counts as a quantitative representation of protein abundance (Lundgren *et al.*, 2010), specifically the exponentially modified protein abundance index (emPAI) since this score considers the number of peptides per protein, an important metric in IP-MS (Ishihama *et al.*, 2005). Additionally, other label free or labeled quantitative values can be used, including results from intensity-based quantification, stable isotope labeling using amino acids in cell culture (SILAC), and tandem mass tag (TMT) experiments (Thompson *et al.*, 2003; Ong and Mann, 2006; Zhang *et al.*, 2010). However, the data must be in an array format and representative of the abundance of a protein. If a protein is not present in either a control or experimental purification, a missing value, CANVS assigns the missing protein half the minimum value of non-missing proteins in the same purification (Wei *et al.*, 2018). CANVS has two methods by which the user can filter data, using the log₂fold change of proteins compared to a control or using a combination of log₂fold change and significance statistics in the form of a two-tailed t-test (Student, 1908; Hubner *et al.*, 2010). Calculating the difference between the

logarithmic mean protein intensities between experimental and control purifications allows users to identify non-specific binders that center around zero (Hubner *et al.*, 2010; Singh *et al.*, 2016). If a sufficient number of replicates (we recommend three biological replicates and two technical replicates) are used in the analysis, the user also has the option of calculating the significance in the log₂fold change using a two-tailed t-test. Comparing the negative log of the P value to the log₂fold change creates a volcano plot where background proteins cluster at zero (Singh *et al.*, 2016). The interactive volcano plot is displayed in the results box of the *Clean* tab (Figure 2 and Supplemental Figures S2-S4). Additionally, the user has the option of deciding to analyze the data by log₂fold change only or log₂foldchange and significance statistics. This is done by changing the drop-down menu under Analysis Method (Supplemental Material – Cleaning Data). The user can filter by a certain fold change of interest or a different P value, however the preset options are set at a log change of 0.6 and a P value of 0.05. Additionally, if the user is concerned about the multiple testing problem and does not wish to use a strict P value cutoff, we included an option to adjust the P value via the Benjamini and Hochberg method (Benjamini *et al.*, 2001). Briefly, adjusting the P value controls the false discovery rate and therefore corrects for the expected amount of false positives among all positives that rejected the null hypothesis (Jafari and Ansari-Pour, 2019). Results appear in the form of a data table in the *Clean* tab, which are referenced throughout the rest of the pipeline.

Alternatively, users can elect to not carry out a statistical analysis and use CANVS solely to contextualize and visualize the results. To do so, users can select the *no analysis* option in the *Clean* tab and the program will consider data in the experimental as the

results. Even when no statistical analysis is performed, the user must press the run button in the *Clean* tab to let CANVS know that the data uploaded in the experimental should be used as the results. This feature allows CANVS to be easily integrated with any other statistical pipeline of choice. The experimental data will then appear in the form of a data table in the *Clean* tab, and this data table will be referenced in the rest of the pipeline (Figure 2 and Supplemental Figures S2-S4).

3. Data analysis

CANVS features two tabs to analyze and generate visual network representations of the results (Supplemental Material – Analysis/Visualization). The first *Analyze/Visualize* tab (Supplemental Figure S5-S7) creates networks for all identified proteins. The *GO Analyze/Visualize* tab (Supplemental Figure S8-S9) features a GO-based (Ashburner *et al.*, 2000; Consortium, 2019) filtering tool that allows users to search the protein results for associations with GO terms of interest. The GO database classifies GO terms into three categories; biological processes (BP), molecular function (MF) and cellular component (CC). CANVS links proteins in the results to associated GO terms. Users can search by keyword and CANVS searches for GO terms that have that keyword. GO terms with the keyword and their respective sub-terms are selected. CANVS then renders two objects: a color-coded network representing how the selected GO terms are related and a data table with all the selected GO terms (Figure 3). Users can then select GO terms of interest in the data table, which instructs CANVS to filter the results for proteins that have the selected GO terms. Alternatively, if users want to filter the results with all of the GO terms in the table, no selection is necessary and the network visuals will reflect proteins

associated with all the GO terms. This feature is particularly helpful if users are interested in specific biological processes, molecular functions and/or cellular components and want to search the results for proteins associated with such GO terms.

Once the user determines whether to include all proteins in the results or selects certain proteins based on GO terms, proteins in the result can be further annotated and contextualized. CANVS integrates two main databases: the Biological General Repository for Interaction Datasets (BioGRID v. 4.3) (Oughtred *et al.*, 2019) and the Comprehensive Resource of Mammalian Protein Complexes (CORUM v. 3.0) (Giurgiu *et al.*, 2019). BioGRID contextualizes results in terms of previously identified protein-protein interactions, whereas, CORUM contextualizes the results in terms of known protein-protein-complex information. For both databases, the user first defines an organism of interest and annotations are performed within the context of that organism. See Supplemental Material for a full list of organisms supported by CANVS, BioGRID, and CORUM. By analyzing the results from both databases, the user can identify known PPIs and complex-PPI/As present in the results.

4. Data visualization

After the results have been annotated (using CORUM or BioGRID) or selected for (using associated GO terms) three types of networks are created; protein-protein network (Figure 4A), CORUM protein-protein-complex network (Figure 4B and Supplemental Figure S10), and a BioGRID protein-protein network (Figure 4C and Supplemental Figure S11). Visual representations are powered by visNetwork, an R package that provides a framework for creating visual networks in an interactive environment (Almende, 2021).

Considering a basic network, the user can interact with the graphs by zooming in on certain sections of a network, selecting certain proteins to highlight interactions, and downloading edited networks as png files. Networks can also be reset and created again, a helpful feature when considering different parameters that might influence which proteins are rendered in the networks. The user also has the option to download Cytoscape (Shannon *et al.*, 2003) network files, which can be used in Cytoscape to re-create networks developed using CANVS. See Supplemental Material for a detailed walkthrough on how to use CANVS network visuals.

CONCLUSION

CANVS is an interactive tool that allows scientists to integrate systems biology databases and create PPI/A networks with biologically relevant results. The simple requirements of the application along with its interactive networks, make CANVS a powerful tool to be used in conjunction with other tools or as a standalone pipeline. The application is particularly useful to researchers studying a POI or sets of proteins and want to contextualize the results of their interactome/association experiments. The integration of BioGRID allows the user to compare the results to a wide set of interactome experiments and create PPI networks based on previous PPI data. Similarly, the integration of CORUM allows users to identify protein complexes within their results, providing context as to how a POI or bait might be related to other proteins in the network. Additionally, by creating networks where specific molecular functions, biological processes or cellular components are prioritized, the user can quickly parse through the results and concentrate on PPI/As that are relevant to the user's scientific question. Ultimately, more databases and other statistical methods can be integrated into the pipeline. It's important to note that protein

interactions and associations should be validated biochemically. As examples, we recently used the analytical framework included in CANVS to study the PPI/A networks of DUSP7, which helped to define its regulation of ERK2 during mitosis (Guo *et al.*, 2021), and also to analyze the PPA networks of core spindle assembly checkpoint proteins (Garcia *et al.*, 2021). Overall, CANVS offers an interactive and easy-to-use solution to study PPI/A data that can be used by laboratories without an established proteomic analysis pipeline.

MATERIALS AND METHODS

Testing CANVS with mass spectrometry data from BioID experiments

To test CANVS, we utilized a previously published mass spectrometry data set from BioID-based experiments of core spindle assembly checkpoint proteins (Garcia *et al.*, 2021). Briefly, BioID2-tagged inducible HeLa stable cell lines were generated for core spindle assembly checkpoint (SAC) proteins (BUB1, BUB3, BUBR1, MAD1L1, MAD2L1). These cell lines were induced to express the BioID2-tagged core SAC proteins, incubated with biotin, and BioID purifications were performed in triplicate for each bait except for BUB3 which was performed in duplicate. A BioID2-tag alone cell line was used as control. The purifications were then analyzed by LC-MS/MS and peptide identification was conducted with Mascot (v2.4; Matrix Science, Boston, MA) against the UniProt human database (October 10, 2018). Search parameters included trypsin digestion allowing up to two missed cleavages, carbamidomethyl on cysteine as a fixed modification, oxidation of methionine as a variable modification, 10-ppm peptide mass tolerance, and 0.02-Da fragment mass tolerance. Peptides that surpassed a cut-off score of 20, considering a

5% false discovery rate, were accepted. From the SAC protein BiID panel, 387 proteins were identified with at least two significant peptides and were further processed using CANVS. To perform a statistical analysis of proteins identified in both the control and experimental, emPAI (Ishihama *et al.*, 2005) scores were considered as quantifiable values. Files compatible with CANVS were then generated by summarizing search results by UniProt accession ID, protein description, quantifiable (emPAI score), file name, and associated bait. This dataset is available at the SourceForge directory (<https://sourceforge.net/projects/canvs/files/>) where users can download it for reference.

Statistical analysis

Given quantitative scores, the significance of proteins shared between the experimental and the control can be determined using log₂fold change values and significance statistics (Hubner *et al.*, 2010; Singh *et al.*, 2016). Proteins are first filtered by the number of replicates each protein is present in, which is a number chosen by the user. Proteins with the appropriate replicate count are further tested for significance. Fold change is calculated by comparing the mean value of a protein across experimental purifications to the same mean value of a protein across control purifications. If the protein is not present in either a control or experimental purification, half of the minimum value of that purification is used (Wei *et al.*, 2018). Fold change is calculated by comparing the mean value of a protein across experimental purifications to the same mean value of a protein across control purifications. CANVS then calculates the log base-2 of the fold change since it is beneficial to represent the distribution around zero, a value that indicates no change for a protein between the experimental and control. If a sufficient number of

replicates (we suggest at least three biological replicates and two technical replicates) are used in the analysis, the user has the option of also calculating the significance in the log2fold change using a two-tailed t-test. Additionally the user can adjust the p value using the Benjamini and Hochberg correction. Briefly, adjusting the P value controls the false discovery rate and therefore corrects for the expected amount of false positives among all positives that rejected the null hypothesis (Jafari and Ansari-Pour, 2019).

Integration of System Biology Databases

Three main system biology databases are integrated in CANVS: the Biological General Repository for Interaction Datasets (BioGRID v. 3.5) (Oughtred *et al.*, 2019), the Comprehensive Resource of Mammalian Protein Complexes (CORUM v. 3.0) (Giurgiu *et al.*, 2019), and Gene Ontology (Ashburner *et al.*, 2000; Consortium, 2019). To incorporate each database, the entire database was downloaded and transformed into an R object that is then referenced in a function. In the case of Gene Ontology, both GO terms and sub-terms were merged by GO term ID. In order to identify data that corresponds to certain proteins across all three databases, UniProt accession numbers are translated to common gene names using R package org.HS.eg.db (Carlson, 2019). The package includes local versions of each database. Updates to databases in the package will be performed every six months.

Development of Shiny Dashboard

CANVS follows a Golem framework, with the motivation of creating an easy to use application with simplified modules (Fay *et al.*, 2021). Modules, functions, and helper

functions were created using the framework outlined in the Golem package. To make the app interactive yet intuitive, a dashboard framework was implemented where the user is able to perform a part of the pipeline in each tab. Visual networks are created using visNetwork, an R package designed to create interactive network visuals (Almende, 2021). Briefly, protein results are formatted to include node information, edge information, and annotations for each node including ID and visual components that can be changed by the user (color, shape, etc.). The networks are then rendered in an interactive Shiny session in the format of a Shiny dashboard. The app was deployed using protocols from DesktopDeployR, a framework for deploying self-contained R-based applications in Windows (<https://github.com/wleepang/DesktopDeployR>). For step-by-step instructions on how to use CANVS refer to the Supplemental Material or the main repository.

Installation

CANVS is packaged as an R Shiny dashboard app but since it includes data from large databases, the application is deployed locally. Currently the local version of the app can only be used in a Windows system. To install CANVS, download the CANVS.zip file in the main repository (<https://sourceforge.net/projects/canvs/files/>). Unzip the file in a location where you want to save the application. Open the folder and right click on the CANVS.bat file, then select create a shortcut. Name the shortcut CANVS and then drag the shortcut to the desktop. Then double click on the new shortcut and CANVS will open in a web browser window. Please make sure that there are no antivirus software blocking application connections to ports in the computer and that firewall does not block use of ports.

ACKNOWLEDGEMENTS

This work was supported by the National Institutes of Health NIGMS grant numbers R35GM139539 and R01GM117475 to J.Z.T., any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Institutes of Health NIGMS. Y.A.G. was supported by the UCLA Tumor Cell Biology Training Program (USHHS Ruth L. Kirschstein Institutional National Research Service Award # T32CA009056). E.F.V. was supported by a grant to The University of California, Los Angeles from the Howard Hughes Medical Institute through the James H. Gilliam Fellowships for Advanced Study Program and by a UCLA Molecular Biology Institute Whitcome Fellowship. I.R. was supported by a NSF Louis Stokes Alliances for Minority Participation Bridge to the Doctorate Fellowship and a Cota Robles Fellowship.

CONFLICTS OF INTEREST

The authors have no conflicts of interest with the contents of this article.

FIGURES

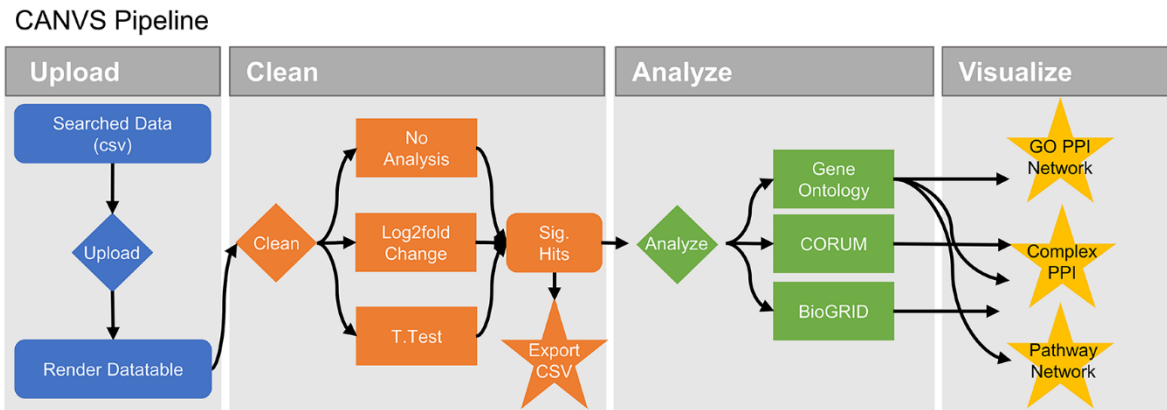


FIGURE 1: CANVS workflow. Mass spectrometry data files, comma-delimited text files, with protein UniProt accession numbers, protein descriptions, protein quantitative values (scores), and bait POIs are uploaded and rendered as interactive data tables. To clean the data, users can determine the significance of the identified proteins, given proper controls, using log2fold change and a t-test. Significant protein identifications are then analyzed by applying Gene Ontology (GO) terms, the Comprehensive Resource of Mammalian Protein Complexes (CORUM) database, and the Biological General Repository for Interaction Datasets (BioGRID) database. The visNetwork R package is then used to visualize the GO PPI/A, CORUM PPI/A, and BIOGRID PPI/A networks.

Analysis

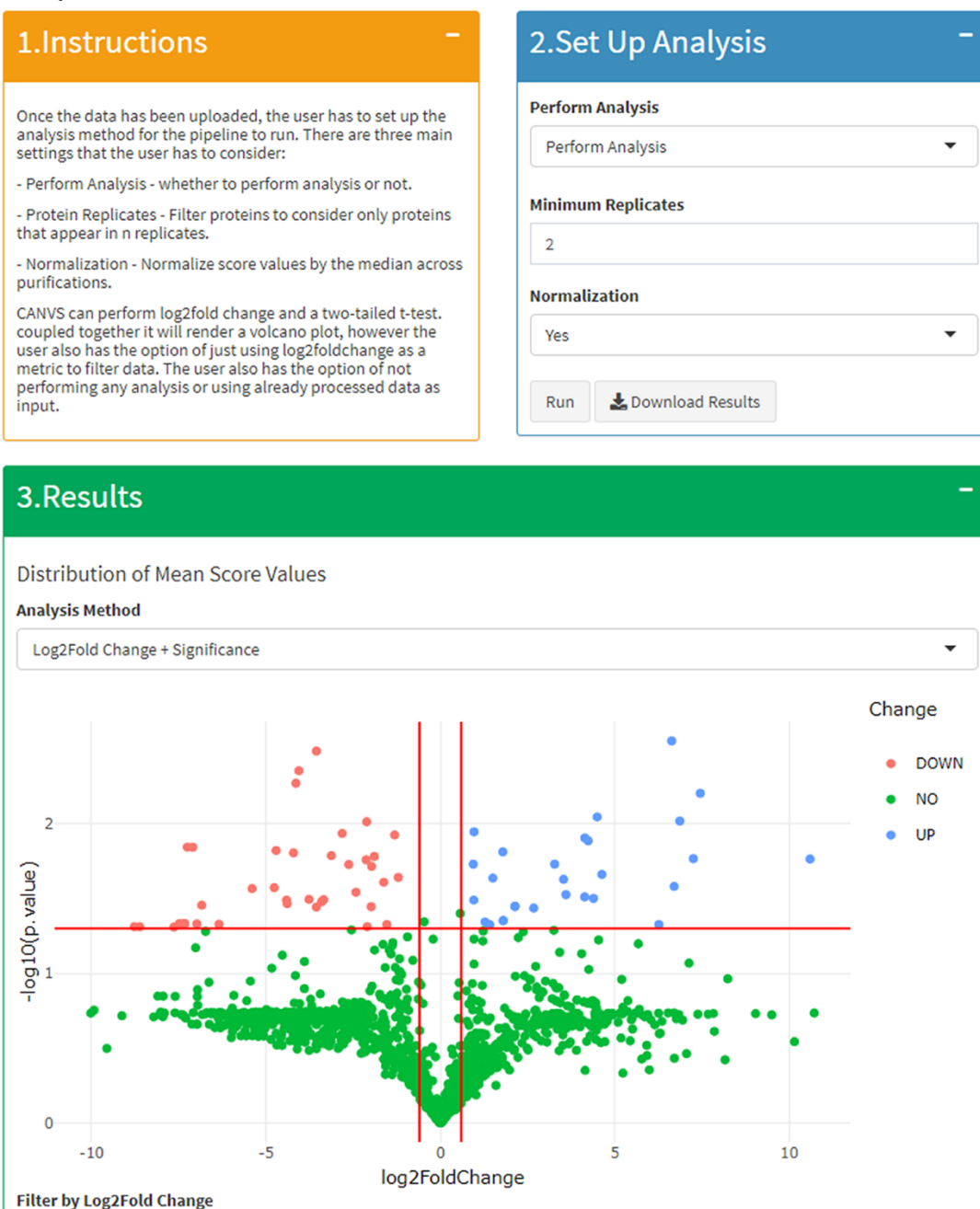


FIGURE 2: CANVS cleaning method. CANVS allows users to upload interaction/association MS data, filter by a minimum number of replicates a protein should be present in, normalize proteins by the median value of each purification, and apply significance statistics. CANVS calculates the log2fold change and P values that can be visualized in an interactive volcano plot. The user can then filter by a certain P value or fold change and the results are used in the pipeline for further analysis/visualization.

Search by Relevant GO Terms

2. Select GO Terms

Search by Keyword

Key:

- Biological Process
- Molecular Function
- Cellular Component

Search Create Network Reset Network Download Cytoscape File

Select by id ▼

Show 10 entries

	GOID	Term	Ontology
4	GO:0090307	mitotic spindle assembly	BP
8	GO:0072686	mitotic spindle	CC
954	GO:1900087	positive regulation of G1/S transition of mitotic cell cycle	BP
1223	GO:0000086	G2/M transition of mitotic cell cycle	BP
1228	GO:0010389	regulation of G2/M transition of mitotic cell cycle	BP
1854	GO:0097431	mitotic spindle pole	CC
1917	GO:1901990	regulation of mitotic cell cycle phase transition	BP
2669	GO:0045930	negative regulation of mitotic cell cycle	BP
2674	GO:0007094	mitotic spindle assembly checkpoint	BP
2675	GO:0007093	mitotic cell cycle checkpoint	BP

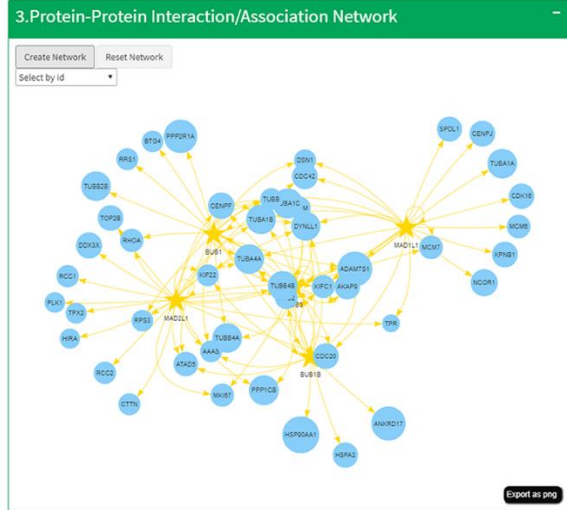
Showing 1 to 10 of 32 entries

Previous 1 2 3 4 Next

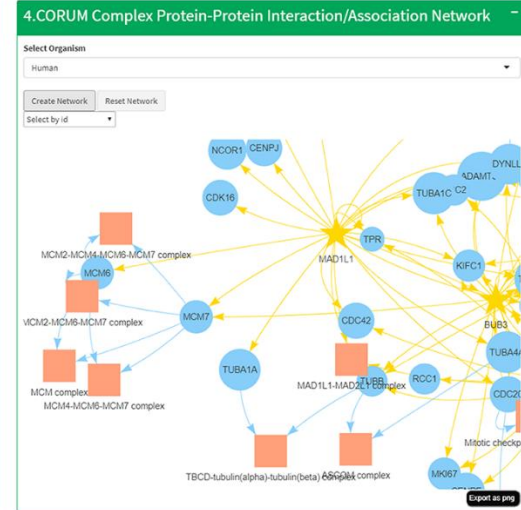
Download GO Terms

FIGURE 3: Selection of Gene Ontology GO terms and filtering results with selected GO terms. Users can perform a keyword search and GO terms containing the keyword/s of interest and any associated sub-terms are retrieved. Only GO terms associated with protein hits in the dataset will appear and can be selected and applied as a filter. Proteins with the associated GO terms of interest are included in the network tables.

A. Create PPI/A Networks



B. Create CORUM PPI/A Networks



C. Create BioGRID PPI/A Networks

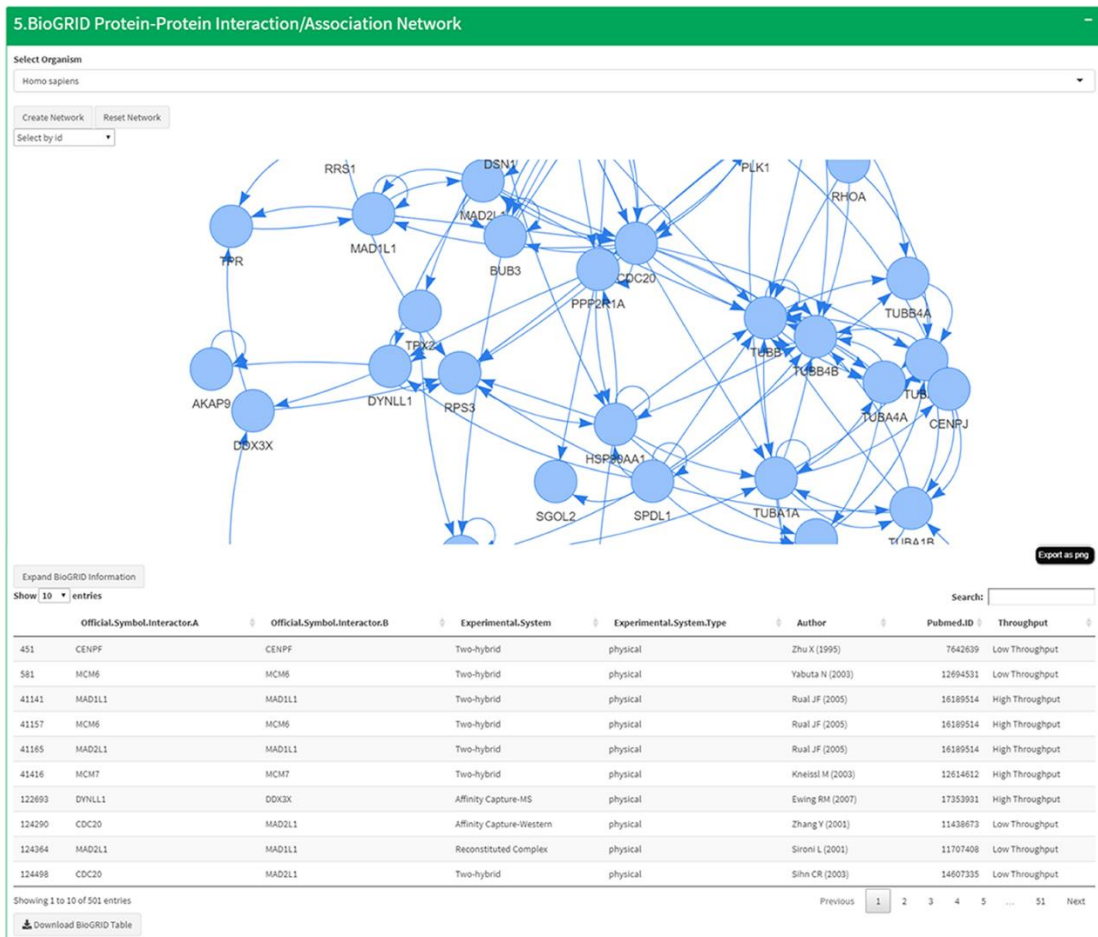


FIGURE 4: Create interactive PPI/A networks of (A) protein hits associated with the selected GO terms, integrating (B) CORUM protein complex information and (C) BioGRID PPI information.

REFERENCES

1. Almende, B.T., Titouan Robert. (2021). CRAN - Package visNetwork.
2. Apweiler, R., Bairoch, A., Wu, C.H., Barker, W.C., Boeckmann, B., Ferro, S., Gasteiger, E., Huang, H., Lopez, R., Magrane, M., Martin, M.J., Natale, D.A., O'Donovan, C., Redaschi, N., and Yeh, L.-S.L. (2004). UniProt: the Universal Protein knowledgebase. *Nucleic acids research* 32.
3. Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., Harris, M.A., Hill, D.P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J.C., Richardson, J.E., Ringwald, M., Rubin, G.M., and Sherlock, G. (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nature genetics* 25.
4. Benjamini, Y., Drai, D., Elmer, G., Kafkafi, N., and Golani, I. (2001). Controlling the false discovery rate in behavior genetics research. *Behavioural brain research* 125.
5. Carlson, M. (2019). org.Hs.eg.db: Genome wide annotation for Human, primarily based on mapping using Entrez Gene identifiers.
6. Chawade, A., Alexandersson, E., and Levander, F. (2014). Normalyzer: a tool for rapid evaluation of normalization methods for omics data sets. *Journal of proteome research* 13.
7. Cheung, K., Senese, S., Kuang, J., Bui, N., Ongpipattanakul, C., Gholkar, A., Cohn, W., Capri, J., Whitelegge, J.P., and Torres, J.Z. (2016). Proteomic Analysis of the Mammalian Katanin Family of Microtubule-severing Enzymes

Defines Katanin p80 subunit B-like 1 (KATNBL1) as a Regulator of Mammalian Katanin Microtubule-severing. *Molecular & cellular proteomics : MCP* 15.

8. Chien, C., Bartel, P., Sternglanz, R., and Fields, S. (1991). The two-hybrid system: a method to identify and clone genes for proteins that interact with a protein of interest. *Proceedings of the National Academy of Sciences of the United States of America* 88.
9. Choi, H., Larsen, B., Lin, Z.-Y., Breitskreutz, A., Mellacheruvu, D., Fermin, D., Qin, Z.S., Tyers, M., Gingras, A.-C., and Nesvizhskii, A.I. (2011). SAINT: probabilistic scoring of affinity purification-mass spectrometry data. *Nature methods* 8.
10. Choi, H., Liu, G., Mellacheruvu, D., Tyers, M., Gingras, A.-C., and Nesvizhskii, A.I. (2012). Analyzing protein-protein interactions from affinity purification-mass spectrometry data with SAINT. *Current protocols in bioinformatics Chapter* 8.
11. Consortium, G.O. (2021). The Gene Ontology resource: enriching a GOld mine. *Nucleic acids research* 49.
12. Consortium, T.G.O. (2019). The Gene Ontology Resource: 20 years and still GOing strong. *Nucleic acids research* 47.
13. Fay, C., Rochette, S., Guyader, V., and Girard, C. (2021). Engineering Production-Grade Shiny Apps.
14. Fields, S., and Song, O. (1989). A novel genetic system to detect protein-protein interactions. *Nature* 340.
15. Firat-Karalar, E.N., and Stearns, T. (2015). Probing mammalian centrosome structure using BioID proximity-dependent biotinylation. *Methods in cell biology* 129.

16. Garcia, Y.A., Velasquez, E.F., Gao, L.W., Gholkar, A.A., Clutario, K.M., Cheung, K., Williams-Hamilton, T., Whitelegge, J.P., and Torres, J.Z. (2021). Mapping Proximity Associations of Core Spindle Assembly Checkpoint Proteins. *Journal of proteome research* 20.
17. Gingras, A., Gstaiger, M., Raught, B., and Aebersold, R. (2007). Analysis of protein complexes using mass spectrometry. *Nature reviews. Molecular cell biology* 8.
18. Giurgiu, M., Reinhard, J., Brauner, B., Dunger-Kaltenbach, I., Fobo, G., Frishman, G., Montrone, C., and Ruepp, A. (2019). CORUM: the comprehensive resource of mammalian protein complexes-2019. *Nucleic acids research* 47.
19. Go, C.D., Knight, J.D.R., Rajasekharan, A., Rathod, B., Hesketh, G.G., Abe, K.T., Youn, J.-Y., Samavarchi-Tehrani, P., Zhang, H., Zhu, L.Y., Popiel, E., Lambert, J.-P., Coyaud, É., Cheung, S.W.T., Rajendran, D., Wong, C.J., Antonicka, H., Pelletier, L., Raught, B., Palazzo, A.F., Shoubridge, E.A., and Gingras, A.-C. (2019). A proximity biotinylation map of a human cell.
20. Guo, X., Ramirez, I., Garcia, Y.A., Velasquez, E.F., Gholkar, A.A., Cohn, W., Whitelegge, J.P., Tofig, B., Damoiseaux, R., and Torres, J.Z. (2021). DUSP7 regulates the activity of ERK2 to promote proper chromosome alignment during cell division. *The Journal of biological chemistry* 296.
21. Hubner, N.C., Bird, A.W., Cox, J., Splettstoesser, B., Bandilla, P., Poser, I., Hyman, A., and Mann, M. (2010). Quantitative proteomics combined with BAC TransgeneOmics reveals in vivo protein interactions. *The Journal of cell biology* 189.

22. Ishihama, Y., Oda, Y., Tabata, T., Sato, T., Nagasu, T., Rappsilber, J., and Mann, M. (2005). Exponentially modified protein abundance index (emPAI) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein. *Molecular & cellular proteomics : MCP* 4.
23. Jafari, M., and Ansari-Pour, N. (2019). Why, When and How to Adjust Your P Values? *Cell journal* 20.
24. Knight, J.D.R., Choi, H., Gupta, G.D., Pelletier, L., Raught, B., Nesvizhskii, A.I., and Gingras, A.-C. (2017). ProHits-viz: a suite of web tools for visualizing interaction proteomics data. *Nature methods* 14.
25. Kuenzi, B.M., Borne, A.L., Li, J., Haura, E.B., Eschrich, S.A., Koomen, J.M., Rix, U., and Stewart, P.A. (2016). APOSTL: An Interactive Galaxy Pipeline for Reproducible Analysis of Affinity Proteomics Data. *Journal of proteome research* 15.
26. Lu, H., Zhou, Q., He, J., Jiang, Z., Peng, C., Tong, R., and Shi, J. (2020). Recent advances in the development of protein-protein interactions modulators: mechanisms and clinical trials. *Signal transduction and targeted therapy* 5.
27. Lundgren, D.H., Hwang, S.-I., Wu, L., and Han, D.K. (2010). Role of spectral counting in quantitative proteomics. *Expert review of proteomics* 7.
28. Michnick, S., Remy, I., Campbell-Valois, F., Vallée-Bélisle, A., and Pelletier, J. (2000). Detection of protein-protein interactions by protein fragment complementation strategies. *Methods in enzymology* 328.

29. Nesvizhskii, A.I. (2012). Computational and informatics strategies for identification of specific protein interaction partners in affinity purification mass spectrometry experiments. *Proteomics* 12.
30. Ong, S.-E., and Mann, M. (2006). A practical recipe for stable isotope labeling by amino acids in cell culture (SILAC). *Nature protocols* 1.
31. Oughtred, R., Stark, C., Breitkreutz, B.-J., Rust, J., Boucher, L., Chang, C., Kolas, N., O'Donnell, L., Leung, G., McAdam, R., Zhang, F., Dolma, S., Willems, A., Coulombe-Huntington, J., Chatr-Aryamontri, A., Dolinsk, K., and Tyers, M. (2019). The BioGRID interaction database: 2019 update. *Nucleic acids research* 47.
32. Perkins, J.R., Diboun, I., Dessailly, B.H., Lees, J.G., and Orengo, C. (2010). Transient protein-protein interactions: structural, functional, and network properties. *Structure (London, England : 1993)* 18.
33. Remnant, L., Booth, D.G., Vargiu, G., Spanos, C., Kerr, A.R.W., and Earnshaw, W.C. (2019). In vitro BioID: mapping the CENP-A microenvironment with high temporal and spatial resolution. *Molecular biology of the cell* 30.
34. Sears, R.M., May, D.G., and Roux, K.J. (2019). BioID as a Tool for Protein-Proximity Labeling in Living Cells. *Methods in molecular biology (Clifton, N.J.)* 2012.
35. Selvin, P. (1995). Fluorescence resonance energy transfer. *Methods in enzymology* 246.

36. Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome research* 13.
37. Singh, S., Hein, M.Y., and Stewart, A.F. (2016). msVolcano: A flexible web application for visualizing quantitative proteomics data. *Proteomics* 16.
38. Student. (1908). The Probable Error of a Mean. *Biometrika* 6, 1-25.
39. Szklarczyk, D., Gable, A.L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., Simonovic, M., Doncheva, N.T., Morris, J.H., Bork, P., Jensen, L.J., and Mering, C.v. (2019). STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic acids research* 47.
40. Teo, G., Liu, G., Zhang, J., Nesvizhskii, A.I., Gingras, A.-C., and Choi, H. (2014). SAINTexpress: improvements and additional features in Significance Analysis of INTeractome software. *Journal of proteomics* 100.
41. Thompson, A., Schäfer, J., Kuhn, K., Kienle, S., Schwarz, J., Schmidt, G., Neumann, T., Johnstone, R., Mohammed, A.K.A., and Hamon, C. (2003). Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. *Analytical chemistry* 75.
42. Torres, J.Z., Summers, M.K., Peterson, D., Brauer, M.J., James Lee, S.S., Gholkar, A.A., Lo, Y.-C., Lei, X., Jung, K., Anderson, D.C., Davis, D.P., Belmont, L., and Jackson, P.K. (2011). The STARD9/Kif16a kinesin associates with mitotic microtubules and regulates spindle pole assembly. *Cell* 147.

43. Välikangas, T., Suomi, T., and Elo, L.L. (2018). A systematic evaluation of normalization methods in quantitative label-free proteomics. *Briefings in bioinformatics* 19.
44. Wei, R., Wang, J., Su, M., Jia, E., Chen, S., Chen, T., and Ni, Y. (2018). Missing Value Imputation Approach for Mass Spectrometry-based Metabolomics Data. *Scientific reports* 8.
45. Yugandhar, K., Gupta, S., and Yu, H. (2019). Inferring Protein-Protein Interaction Networks From Mass Spectrometry-Based Proteomic Approaches: A Mini-Review. *Computational and structural biotechnology journal* 17.
46. Zhang, G., Ueberheide, B.M., Waldemarson, S., Myung, S., Molloy, K., Eriksson, J., Chait, B.T., Neubert, T.A., and Fenyö, D. (2010). Protein quantitation using mass spectrometry. *Methods in molecular biology (Clifton, N.J.)* 673.

CHAPTER 3

Mapping Proximity Associations of Spindle Assembly Checkpoint Proteins

Mapping Proximity Associations of Spindle Assembly Checkpoint Proteins

Yenni A. Garcia¹, Erick F. Velasquez¹, Lucy W. Gao², Ankur A. Gholkar¹, Kevin M. Clutario¹, Keith Cheung¹, Taylor Williams-Hamilton¹, Julian P. Whitelegge^{2,3,4} and Jorge Z. Torres^{1, 3, 4*}

¹Department of Chemistry and Biochemistry, University of California, Los Angeles, CA 90095, USA

²Pasarow Mass Spectrometry Laboratory, The Jane and Terry Semel Institute for Neuroscience and Human Behavior, David Geffen School of Medicine, University of California, Los Angeles, CA 90095, USA

³Molecular Biology Institute, University of California, Los Angeles, CA 90095, USA

⁴Jonsson Comprehensive Cancer Center, University of California, Los Angeles, CA 90095, USA

*Corresponding author:

Jorge Z. Torres
607 Charles E. Young Drive East
Los Angeles, CA 90095
Phone: 310-206-2092
Fax: 310-206-5213
torres@chem.ucla.edu

Reprinted (adapted) with permission from {Garcia YA, Velasquez EF, Gao LW, Gholkar AA, Clutario KM, Cheung K, Williams-Hamilton T, Whitelegge JP, Torres JZ. Mapping Proximity Associations of Core Spindle Assembly Checkpoint Proteins. J Proteome Res. 2021 Jul 2;20(7):3414-3427. doi: 10.1021/acs.jproteome.0c00941. Epub 2021 Jun 4. PMID: 34087075; PMCID: PMC8256817.}. Copyright {2021} American Chemical Society.

Abbreviations: BiO2, Biotin identification 2; SAC, Spindle assembly checkpoint

ABSTRACT

The spindle assembly checkpoint (SAC) is critical for sensing defective microtubule-kinetochore attachments and tension across the kinetochore and functions to arrest cells in prometaphase to allow time to repair any errors before proceeding into anaphase. Dysregulation of the SAC leads to chromosome segregation errors that have been linked to human diseases like cancer. Although much has been learned about the composition of the SAC and the factors that regulate its activity, the proximity associations of core SAC components have not been explored in a systematic manner. Here, we've taken a BioID2 proximity-labeling proteomic approach to define the proximity protein environment for each of the five core SAC proteins BUB1, BUB3, BUBR1, MAD1L1, and MAD2L1 in mitotic-enriched populations of cells where the SAC is active. These five protein association maps were integrated to generate a SAC proximity protein network that contains multiple layers of information related to core SAC protein complexes, protein-protein interactions, and proximity associations. Our analysis validated many known SAC complexes and protein-protein interactions. Additionally, it uncovered new protein associations, including the ELYS-MAD1L1 interaction that we have validated, that lend insight into the functioning of core SAC proteins and highlight future areas of investigation to better understand the SAC.

Key words: Spindle assembly checkpoint (SAC), BioID2, Proximity labeling, Protein associations, Protein networks, Cell division

INTRODUCTION

Human cell division is a highly coordinated set of events that ensures the proper transmission of genetic material from one mother cell to two newly formed daughter cells. Chromosome missegregation during cell division can lead to aneuploidy, an aberrant chromosomal number, which is a hallmark of many types of cancers and has been proposed to promote tumorigenesis (1). However, there is currently no consensus as to the pathways and factors that are deregulated to induce aneuploidy, why it is prevalent in cancer and how it contributes to tumorigenesis. Pivotal to cell division is the metaphase to anaphase transition, which is a particularly regulated process involving a multitude of protein-protein interactions that relies heavily on posttranslational modifications like phosphorylation and ubiquitination that function as switches to activate or inactivate protein function (2,3). For example, the multi-component spindle assembly checkpoint (SAC) is activated when unattached kinetochores or nonproductive (monotelic, syntelic, and merotelic) attachments are sensed and functions to arrest cells in metaphase to give time to correct these deficiencies and generate proper microtubule-kinetochore attachments (2) (Figure 1A). This ensures proper sister chromatid separation and minimizes segregation errors that lead to chromosomal instability, aneuploidy, and tumorigenesis (1). Core components of the SAC include BUB1, BUB3, BUBR1, MAD1L1, and MAD2L1(4). Critical to the SAC is the mitotic checkpoint complex (MCC, composed of MAD2L1, BUBR1, BUB3, and CDC20) that maintains the anaphase promoting complex/cyclosome (APC/C) ubiquitin ligase substrate adaptor protein CDC20 sequestered and thereby inactivates the APC/C (5,6). Upon proper microtubule-kinetochore attachment the SAC is satisfied and the inhibitory effect of the MCC on the

APC/C is relieved (2) (Figure 1A). Active APC/C then ubiquitinates and targets Securin for degradation (2), which activates Separase, the protease that cleaves RAD21, a component of the cohesin complex that holds sister chromatids together (7). This releases sister chromatid cohesion and chromatids are pulled to opposing poles of the cell by spindle microtubules, marking the entry into anaphase.

Because understanding the SAC is critical to understanding tumorigenesis and the response of tumor cells to antimitotic drugs that activate the SAC and trigger apoptotic cell death, it has become an intensive area of research (8,9). Although decades of research have shed light on the SAC, we are far from elucidating the full complement of regulatory factors involved in this complex pathway and from understanding how misregulation of this pathway can lead to tumorigenesis and resistance to chemotherapeutic drugs like antimitotics (10). Furthermore, models of proximity associations of the core SAC proteins with themselves and with structural and signaling components that mediate the establishment and silencing of the SAC are still being defined (11-13). Recently, proximity-labeling approaches like BioID and APEX have been used effectively to determine association networks among proteins and for defining the architecture of the centrosome, centrosome-cilia interface, and other organelles within the cell (14-19). However, proximity labeling has not been applied to the SAC in a systematic fashion, which could help to interrogate current models of core SAC protein associations and regulation.

Here, we have engineered vectors for establishing inducible BioID2-tagged protein stable cell lines. This system was used to establish stable cell lines with inducible BioID2-tagged core SAC protein (BUB1, BUB3, BUBR1, MAD1L1, and MAD2L1) expression.

These cell lines were utilized in BioID2-proximity biotin labeling studies, which were coupled to biotin biochemical purifications and mass spectrometry analyses to map the associations among the core SAC proteins and other proteins in close proximity. These analyses yielded a wealth of information with regards to the protein environment of the core SAC proteins in mitotic-enriched populations of cells where the SAC is active. In addition to validating well-established SAC protein complexes and protein-protein interactions, we defined new protein associations that warrant further investigation, including the ELYS-MAD1L1 interaction, to advance our understand SAC protein function and regulation.

EXPERIMENTAL PROCEDURES

Cell Culture and Cell Cycle Synchronization

All media and chemicals were purchased from ThermoFisher Scientific (Waltham, MA) unless otherwise noted. HeLa Flp-In T-REx BioID2-tagged stable cell lines and RPE cells were grown in F12:DMEM 50:50 medium with 10% FBS, 2 mM L-glutamine, in 5% CO₂ at 37° C. Cells were induced to express the indicated BioID2-tagged proteins by the addition of 0.2 µg/ml doxycycline (Sigma-Aldrich, St. Louis, MO) for 16 hours. For synchronization of cells in mitosis, cells were treated with 100 nM Taxol (Sigma-Aldrich) for 16 hours. A list of all reagents used is provided in Table S1.

Cell siRNA and Chemical Treatments

HeLa cell siRNA treatments were performed as described previously (20), with control siRNA (siControl, D-001810-10) or BUB1-targeting siRNA (siBUB1, L-004102-00) from

Dharmacon (Lafayette, CO) for 48 hours. For chemical treatments, RPE or HeLa cells were treated with control DMSO vehicle or the BUB1 inhibitor BAY 1816032 (HY-103020) (21) from MedChemExpress (Monmouth Junction, NJ) at 10 μ M for five hours.

Generation of Inducible BioID2-tagging Vectors and Stable Cell Lines

For generating pGBioID2-27 or pGBioID2-47 vectors, the EGFP-S-tag was removed from pGLAP1 (22) by digestion with BstBI and AflIII. BioID2-Myc-27 (27 amino acid linker) or BioID2-Myc-47 (47 amino acid linker) were PCR amplified, digested with NheI and XhoI and cloned into BstBI and AflIII digested pGLAP1 to generate pGBioID2-27 or pGBioID2-47 (Figure S1A). For full-length human SAC core gene *hBUB1*, *hBUB3*, *hBUBR1*, *hMAD1L1*, and *hMAD2L1* expression, cDNA corresponding to the full-length open reading frame of each gene was cloned into pDONR221 as described previously (22,23) (Figure S1B). SAC core genes were then transferred from pDONR221 to pGBioID2-47 using the Gateway cloning system (Invitrogen, Carlsbad, CA) as described previously (22,23) (Figure S1B). The pGBioID2-47-SAC protein vectors were then used to generate doxycycline inducible HeLa Flp-In T-REx BioID2 stable cell lines that expressed the fusion proteins from a specific single locus within the genome as described previously (22,23) (Figure S1C,D). All primers were purchased from ThermoFisher Scientific. A list of primers used is provided in Table S2. For a list of vectors generated in this study see Table S3. The pGBioID2-27 and pGBioID2-47 vectors have been deposited at Addgene (AddgeneIDs: 140276 and 140277 respectively) and are available to the scientific community.

Biotin Affinity Purifications

All media, chemicals, and beads were purchased from ThermoFisher Scientific unless otherwise noted. Biotin affinity purifications were conducted using previously described protocols with modifications (18,19). Briefly, 10% FBS was treated with 1 ml of MyOne streptavidin C1 Dynabeads overnight and passed through a 0.22 μm filter. The BioID2-BUB1, BUB3, BUBR1, MAD1L1, and MAD2L1, and BioID2 alone inducible stable cell lines were plated on six 150 mm tissue culture dishes, 24 hours post-plating, the cells were washed three times with PBS and once with DMEM without FBS, and shifted to the streptavidin Dynabead-treated 10% FBS DMEM. The cells were induced with 0.2 $\mu\text{g/ml}$ Dox, and treated with 100 nM Taxol and 50 μM Biotin for 16 hours. Mitotic cells were collected by shake-off and centrifuged at 1,500 rpm for 5 minutes and washed twice with PBS. The pellet was lysed with 3 ml of lysis buffer (50 mM Tris-HCl pH 7.5, 150 mM NaCl, 1 mM EDTA, 1 mM EGTA, 1% Triton-X-100, 0.1% SDS, Halt Protease and Phosphatase Inhibitor Cocktail) and incubated with gentle rotation for 1 hour at 4 ° C, then centrifuged at 15,000 rpm for 15 minutes and transferred to a new 15 ml conical tube. The lysate was transferred to a TLA-100.3 tube (Beckman Coulter, Indianapolis, IN) and centrifuged at 45,000 rpm for 1 hour at 4 ° C. The lysate was then transferred to a new 15 ml conical tube and incubated with 300 ml of equilibrated streptavidin Dynabeads overnight with gentle rotation at 4 ° C. The beads were separated with a magnetic stand and washed twice with 2% SDS, followed by a wash with WB1 (0.1% sodium deoxycholate, 1% Triton X-100, 500 mM NaCl, 1 mM EDTA, 50 mM HEPES), a wash with WB2 (250 mM LiCl, 0.5% deoxycholate, 1 mM EDTA, 10 mM Tris-HCl pH 8.0), and a final wash with 50 mM Tris-HCl pH 7.5. The beads were then resuspended in 50 mM triethylammonium

bicarbonate (TEAB), 12 mM sodium lauroyl sarcosine, 0.5% sodium deoxycholate. 10% of the beads were boiled with sample buffer and used for immunoblot analysis.

In Solution Tryptic Digestion

Streptavidin Dynabeads in 50 mM triethylammonium bicarbonate (TEAB), 12 mM sodium lauroyl sarcosine, 0.5% sodium deoxycholate were heated to 95 ° C for 10 minutes and then sonicated for 10 minutes to denature proteins. Protein disulfide bonds were reduced by treatment with 5 mM tris(2-carboxyethyl) phosphine (final concentration) for 30 minutes at 37° C. Protein alkylation was performed with 10 mM chloroacetamide (final concentration) and incubation in the dark for 30 minutes at room temperature. The protein solutions were diluted five-fold with 50 mM TEAB. Trypsin was prepared in 50 mM TEAB and added 1:100 (mass:mass) ratio to target proteins followed by a 4-hour incubation at 37 ° C. Trypsin was again prepared in 50 mM TEAB and added 1:100 (mass:mass) ratio to target proteins followed by overnight incubation at 37 ° C. A 1:1 (volume:volume) ratio of ethyl acetate plus 1% trifluoroacetic acid (TFA) was added to the samples and samples were vortexed for five minutes. Samples were centrifuged at 16,000 x g for five minutes at room temperature and the supernatant was discarded. Samples were then lyophilized by SpeedVac (ThermoFisher Scientific) and desalted on C18 StageTips (ThermoFisher Scientific) as described previously (24).

Nano-liquid Chromatography with Tandem Mass Spectrometry (LC-MS/MS)

Analysis Nano-LC-MS/MS with collision-induced dissociation was performed on a Q Exactive Plus Orbitrap (ThermoFisher Scientific) integrated with an Eksigent 2D nano-LC

instrument. A laser-pulled reverse-phase column, 75 μm x 200 mm, containing 5- μm C18 resin with 300-Å pores (ThermoFisher Scientific) was used for online peptide chromatography. Electrospray ionization conditions using the nanospray source (ThermoFisher Scientific) for the Orbitrap were set as follows: capillary temperature at 200° C, tube lens at 110 V, and spray voltage at 2.3 kV. The flow rate for reverse-phase chromatography was 500 nl/min for loading and analytical separation (buffer A, 0.1% formic acid and 2% acetonitrile; buffer B, 0.1% formic acid and 98% acetonitrile). Peptides were loaded onto the column for 30 minutes and resolved by a gradient of 0–80% buffer B over 174 minutes. The Q Exactive Plus Orbitrap was operated in data-dependent mode with a full precursor scan time at 180 minutes at high resolution (70,000 at m/z 400) from 350-1,700 m/z and 10 MS/MS fragmentation scans at low resolution in the linear trap using charge-state screening excluding both unassigned and +1 charge ions. For collision-induced dissociation, the intensity threshold was set to 500 counts, and a collision energy of 40% was applied. Dynamic exclusion was set with a repeat count of 1 and exclusion duration of 15 seconds.

Experimental Design and Statistical Rationale

To enhance confidence in identifying core SAC protein proximity associations, we performed control and experimental purifications in biological replicates (3 biological purifications for each core SAC proteins, except for BUB3 where 2 biological purifications were performed, and 2 technical replicates were performed for each biological purification). This approach allowed for downstream comparison of control and experimental purifications, where proteins identified in the control BirA only (empty vector)

were deemed potential non-specific associations. See Figure S2 for experimental mass spectrometry data acquisition and analysis workflow. Database searches of the acquired spectra were analyzed with Mascot (v2.4; Matrix Science, Boston, MA) as described previously (25). The UniProt human database (October 10, 2018) was used with the following search parameters: trypsin digestion allowing up to 2 missed cleavages, carbamidomethyl on cysteine as a fixed modification, oxidation of methionine as a variable modification, 10-ppm peptide mass tolerance, and 0.02-Da fragment mass tolerance. With these parameters, an overall 5% peptide false discovery rate, which accounts for total false positives and false negatives, was obtained using the reverse UniProt human database as the decoy database. Peptides that surpassed an expectation cut-off score of 20 were accepted. See Table S4 for a list of all identified peptides and Table S5 for a list of all identified proteins. A list of all peptides that were used to identify proteins with one peptide sequence is provided in Table S6. All raw mass spectrometry files can be accessed at the UCSD Center for Computational Mass Spectrometry MassIVE datasets <ftp://MSV000084975@massive.ucsd.edu>. Peptides meeting the above criteria with information about their corresponding identified protein were further analyzed using in-house R scripts. All R scripts used in this study are freely available at GitHub <https://github.com/uclatorreslab/MassSpecAnalysis>. To increase precision and reduce error, a pseudo qualitative/quantitative approach was taken. Proteins identified in both the control and test purifications were assayed for significance, whereas proteins identified in test purifications but not present in control purifications were further considered. To handle proteins shared between test and control purifications, but only identified in less frequency, we measured the relative fold change or mean difference in

a quantitative manner. To compare quantification between purifications, we used the Exponentially Modified Protein Abundance Index (emPAI) (26). emPAI offers approximate relative quantitation of the proteins in a mixture based on protein coverage by the peptide matches in a database search result and can be calculated using the following equation (26).

$$emPAI = 10^{\frac{N_{Observed}}{N_{Observable}}} - 1$$

Where $N_{Observed}$ is the number of experimentally observed peptides and $N_{Observable}$ is the calculated number of observable peptides for each protein (26). To compare proteins across multiple replicates/baits each emPAI score was normalized to pyruvate carboxylase, a protein that readily binds to biotin (27), and was found in high abundance in all purifications. Using a normalized emPAI (NemPAI) as a relative quantification score, we calculated the mean difference (the mean NemPAI for a certain protein across test replicates minus the mean NemPAI). Resampling involved recreating or estimating the normal distribution around a test statistic, in this case the mean difference, by calculating that statistic many times under rearrangement of labels. We performed ten thousand simulations per test statistic, resulting in normal distributions of mean difference between values of proteins identified in the experimental and the control. Using this distribution, we related each individual mean difference to the mean difference observed in the overall population in order to get a relative idea of what might be significantly higher in value compared to the control, when taking what is observed in the entire population. Values that lied outside of the 95% confidence interval of the mean difference and showed a higher value in the experimental compared to the control were then considered for further analysis (see Table S7).

Protein Proximity Network Visualization and Integration of Systems Biology Databases

Visual renderings relating protein-protein interactions/associations were carried out using custom scripts in R. To incorporate protein-complex information, we integrated the Comprehensive Resource of Mammalian Protein Complexes (CORUM v. 3.0) (28). Protein-protein interaction information was derived and integrated from the Biological General Repository for Interaction Datasets (BioGRID v. 3.5) (29). To create relational networks that associated proteins based on cellular mechanisms, Gene Ontology (GO) terms were incorporated into the search (Gene Ontology release June 2019) (30). For a list of GO terms used, see Table S8. Pathway information was derived from Reactome, an open source and peer-reviewed pathway database (31). All databases were individually curated into an in-house systems biology relational database using custom R scripts. Final visuals relating protein associations were constructed using RCytoscapeJS, a developmental tool used to develop Cytoscape renderings in an R and JavaScript environment (32,33).

Immunoprecipitations

For cell lysate immunoprecipitations (IPs), BioID2 (empty vector, EV), BioID2-MAD1L1, or BioID2-MAD2L1 HeLa stable cell lines were induced with 0.2 $\mu\text{g/ml}$ Dox and treated with 100 nM Taxol for 16 hours to arrest cells in mitosis. Cells were collected by shake-off and lysed with lysis buffer (50 mM Tris-HCl pH 7.5, 150 mM NaCl, 1 mM EDTA, 1 mM EGTA, 1% Triton-X-100, 0.1% SDS, Halt Protease and Phosphatase Inhibitor Cocktail)

and incubated with gentle rotation for one hour at 4 ° C, then centrifuged at 15,000 rpm for 30 minutes and the supernatant was transferred to a microcentrifuge tube. Myc magnetic beads were equilibrated and incubated with mitotic cell extracts for five hours at 4 ° C with gentle rotation. The beads were then washed five times with wash buffer (50 mM Tris pH 7.4, 150 mM NaCl, 1 mM DTT, and Halt Protease and Phosphatase Inhibitor Cocktail) for five minutes each and bound proteins were eluted with 50 µL of 2X Laemmli SDS sample buffer. Ten percent of the sample inputs, and the entire eluates from the immunoprecipitations were used for immunoblot analysis.

In Vitro Binding Assays

For *in vitro* binding assays, Myc or FLAG-tagged GFP, MAD1L1, MAD2L1, or ELYS (N-terminal fragment) were *in vitro* transcribed and translated (IVT) using TNT® Quick Coupled Transcription/Translation System, (Promega, Madison, WI) in 10 µL reactions. Myc beads (MBL, Sunnyvale, CA) were washed three times and equilibrated with wash buffer (50 mM Tris pH 7.4, 200 mM KCl, 1 mM DTT, 0.5% NP-40, and Halt Protease and Phosphatase Inhibitor Cocktail). IVT reactions were added to the equilibrated Myc beads and incubated for 1.5 hours at 30 ° C with gentle shaking and after binding, beads were washed three times with wash buffer and eluted by boiling for 10 minutes with 2X Laemmli SDS sample buffer. The samples were then resolved using a 4-20% gradient Tris gel with Tris-Glycine SDS running buffer, transferred to an Immobilon PVDF membrane (EMD Millipore, Burlington, MA), and the membranes were analyzed using a PharosFX Plus molecular imaging system (Bio-Rad, Hercules, CA).

Immunofluorescence Microscopy

Immunofluorescence microscopy was performed as described previously (34) with modifications described in (25). Briefly, HeLa inducible BioID2-tagged BUB1, BUB3, BUBR1, MAD1L1, and MAD2L1 stable cell lines were treated with 0.2 $\mu\text{g/ml}$ doxycycline for 16 hours, fixed with 4% paraformaldehyde, permeabilized with 0.2% Triton X-100/PBS, and co-stained with 0.5 $\mu\text{g/ml}$ Hoechst 33342 and the indicated antibodies. Imaging of mitotic cells was carried out with a Leica DMI6000 microscope (Leica DFC360 FX Camera, 63x/1.40-0.60 NA oil objective, Leica AF6000 software, Buffalo Grove, IL) at room temperature. Images were subjected to Leica Application Suite 3D Deconvolution software and exported as TIFF files. The quantification of immunofluorescence microscopy images from BUB1 RNAi and BUB1 inhibitor treated cells was performed by capturing intensity profiles in ImageJ for both a kinetochore section and a background section adjacent to the kinetochore. Each intensity value was normalized by the area of the captured image and the background signal was subtracted. The values were compared using a student's t-test. The number of samples used varied by experiment; knock-down experiments: BUB1 (n=19), SGO2 (n=50), and PLK1 (n=13); inhibitor treatments: BUB1 (n=20), SGO2 (n=17), and PLK1 (n=17). All calculations were performed in R.

Antibodies

Immunofluorescence microscopy and immunoblotting were performed using the following antibodies: BioID2 (BioFront Technologies, Tallahassee, FL), GAPDH (Preteintech, Rosemont, IL), α -tubulin (Serotec, Raleigh, NC), anti-centromere antibody (ACA, Cortex

Biochem, Concord, MA), SGO2 (Bethyl, Montgomery, TX), PLK1, BUB1, and ELYS (Abcam, Cambridge, MA). Affinipure secondary antibodies labeled with FITC, Cy3, and Cy5 were purchased from Jackson Immuno Research (West Grove, PA). IRDye 680RD streptavidin was purchased from LI-COR Biosciences (Lincoln, NE). Immunoblot analyses were carried out using secondary antibodies conjugated to IRDye 680 and IRDye 800 from LI-COR Biosciences (Lincoln, NE) and blots were scanned using a LI-COR Odyssey infrared imager.

RESULTS AND DISCUSSION

Generation of Inducible BioID2-tagged SAC Protein Stable Cell Lines

The spindle assembly checkpoint is essential for ensuring the fidelity of chromosome segregation during cell division (35) (Figure 1A). To better understand how the SAC functions and is regulated, we sought to map the protein associations of the core SAC proteins BUB1, BUB3, BUBR1 (BUB1B), MAD1L1, and MAD2L1 using a BioID2 proximity labeling proteomic approach (18) (Figure 1B-F). The over-expression of critical cell division proteins often leads to cell division defects that can preclude the generation of epitope-tagged stable cell lines. Therefore, we first sought to generate BioID2 Gateway-compatible vectors with a doxycycline (Dox) inducible expression functionality. To do this, we amplified BirA-Myc with linkers coding for 27 or 47 amino acid residues downstream of Myc (BirA-Myc-27/47) (Figure S1A, Table S2). These amplification products were cloned into the pGLAP1 vector (22), which had been previously modified by removal of its LAP-tag (EGFP-Tev-S-protein), to generate the pGBioID2-27 and pGBioID2-47 vectors (Figure S1A). Full-length human open reading frames encoding for BUB1, BUB3,

BUBR1, MAD1L1, and MAD2L1 were cloned into the pGBioID2-47 vector. The pGBioID2-47-SAC protein vectors (Figure S1B, Table S3), were co-transfected with a vector expressing the Flp recombinase (pOG44) into HeLa Flp-In T-REx cells (Figure S1C). Hygromycin resistant clones were then selected (Figure S1D) and grown in the presence or absence of Dox for 16 hours. The Dox-induced expression of each BioID2-47-SAC protein was then assessed by immunoblot analysis (Figure 2A). All of the BioID2-tagged core SAC proteins were expressed only in the presence Dox (Figure 2A), indicating the successful establishment of inducible BioID2-tagged core SAC protein stable cell lines. Additionally, these BioID2-tagged core SAC proteins were expressed at lower levels than the untagged endogenous proteins (Figure S3A)

BioID2-SAC Proteins Localize Properly to Kinetochores During Prometaphase

Next the ability of BioID2-SAC proteins to properly localize to the kinetochores during prometaphase, a time when the SAC is active and core SAC proteins localize to the kinetochore region, was analyzed by immunofluorescence microscopy. BioID2-SAC protein HeLa inducible stable cells lines were treated with Dox for 16 hours, fixed, and stained with Hoechst 33342 DNA dye and anti-BioID2, anti- α -Tubulin and anti-centromere antibodies (ACA). The localization of BioID2-SAC proteins in prometaphase cells was then monitored by immunofluorescence microscopy. BioID2-tagged BUB1, BUB3, BUBR1, MAD1L1, and MAD2L1 localized to kinetochores, overlapping fluorescence signal with anti-centromere antibodies (ACA) during prometaphase (Figure 2B). In contrast, the BioID2-tag alone showed no specific localization (Figure 2B). These results indicated that the BioID2-tag was not perturbing the ability of the SAC proteins to localize

to kinetochores during the time when the SAC was active. Further, the addition of biotin did not perturb the localization of the BioID2-SAC proteins to the kinetochores (Figure S3B).

BioID2-SAC Protein Proximity Labeling, Purifications, and Peptide Identification

To define the protein proximity networks of core SAC proteins, the inducible BioID2-SAC protein HeLa stable cell lines were used to perform BioID2-dependent proximity biotin labeling and biotinylated proteins were purified with a streptavidin resin (Figures 1D and 2C). Briefly, inducible BioID2-SAC protein HeLa stable cells lines were treated with 0.2 μ g/ml Dox, 100 nM Taxol, and 50 μ M Biotin for 16 hours to induce the expression of BioID2-SAC proteins and to activate the SAC and arrest cells in prometaphase. Mitotic cells were collected by shake-off, lysed, and the cleared lysates were bound to streptavidin beads. Bound biotinylated proteins were trypsinized on the beads and the peptides were analyzed by 2D-LC MS/MS (for details see Experimental Procedures). A diagnostic immunoblot analysis of each purification, using anti-BioID2 antibodies, showed that BioID2-tagged BUB1, BUB3, BUBR1, MAD1L1, and MAD2L1 were present in the extracts and were purified with the streptavidin beads, indicating that they had been biotinylated (Figure 2C). Additionally, western blots of each purification were probed with streptavidin, which showed that biotinylated proteins were present and efficiently captured in each purification (Figure S4A). In-house R scripts were then used to analyze the mass spectrometry results (for details see Experimental Procedures), to draw significance between peptides shared between the experimental and the control, we estimated the distribution of the mean difference of normalized emPAI scores across proteins and

selected proteins with a significant higher difference (for details see Experimental Procedures). Proteins that showed significant higher values in test purifications compared to the controls (values that lied outside of 95% confidence interval of the population mean difference) were considered hits and further analyzed (Table S7).

Analysis of the Core SAC Protein Proximity Association Network

In-house R scripts were then used to integrate the identified proteins from the mass spectrometry analysis with the data visualization application RCytoscapeJS (32) to generate protein proximity association maps for each of the core SAC proteins (BUB1; BUB3; BUBR1; MAD1L1; MAD2L1) (Figure S5). These five maps were compiled to generate the SAC protein proximity network (Figure S6). To begin to digest the wealth of information within the SAC protein proximity network, we first analyzed the network with the CORUM database (28) and examined the proximal associations between each of the core SAC proteins. This analysis revealed many of the previously characterized core SAC component protein-protein interactions and the BUB1-BUB3, BUBR1-BUB3, BUBR1-BUB3-CDC20 (BBC subcomplex of the MCC) and MAD2L1-BUBR1-BUB3-CDC20 (MCC) complexes (Figures 3 and S6) (6,36-38). These SAC complexes are critical to the establishment and maintenance of the SAC (39) and their identification was an indication that our proximity-based labeling approach was robust. Of interest, BUB3 was present in all of the purifications, consistent with its central role in recruiting other SAC proteins to the kinetochore and coordinating the formation of SAC sub-complexes (Figure 3) (12). Although MAD1L1 and MAD2L1 had been previously determined to bind directly (40), our approach was unable to detect this association. However, previous proteomic analyses

with N- or C- terminal BioID-tagged MAD1L1 were also unable to detect an association with MAD2L1, which was attributed to a low number of lysines on the surface of MAD2L1 that likely affected the efficiency of biotin labeling (41).

Analysis of Core SAC Protein-Kinetochores Protein Proximity Associations

To specifically analyze the kinetochores proteins identified in the core SAC protein proximity networks, we applied a kinetochores related Gene Ontology (GO) annotation analysis to the data sets. Briefly, R scripts were used to integrate the identified proteins with the bioinformatic databases CORUM (28), Gene Ontology (30), BioGRID (29), and Reactome (31) using kinetochores related GO terms (see Table S8 for a list of Kinetochores GO IDs) to reveal the kinetochores associated proteins. RCytoscapeJS (32) was then used to generate GO, BioGRID, and Reactome kinetochores protein proximity association maps for each of the core SAC proteins (BUB1; BUB3; BUBR1; MAD1L1; MAD2L1) (Figures S7-S11). The five kinetochores GO maps (one for each core SAC protein) were compiled to generate one core SAC protein kinetochores GO network that visualized the proteins within the network that were active at the kinetochores (Figure S12A). A similar process was repeated to generate one core SAC protein BioGRID network that displayed the verified associations between the proteins that were active at the kinetochores (Figure S12B) and one core SAC protein Reactome network that highlighted the cellular pathways that proteins in the SAC proximity association network have been linked to (Figure S12C). Additionally, we generated core SAC protein GO, BioGRID, and Reactome networks using mitotic spindle related GO annotations (Figure S13A-C) and centromere related GO annotations (Figure S14A-C), see Table S8 for a list of GO IDs.

Finally, we generated core SAC protein GO, BioGRID, and Reactome networks using the kinetochore, mitotic spindle, and centromere related GO annotations (Figure 4A-C). Interestingly, of the proteins identified in the purifications, kinetochore associated proteins were enriched in comparison to mitochondrial proteins (Figure S15). Together, these networks not only visualized the associations of each core SAC protein with kinetochore components and more broadly proteins implicated in mitotic spindle assembly, they also provided a holistic view of their interconnectedness (ie. associations among core SAC proteins and subcomplex and complex formation).

Numerous insights were derived from these networks and we highlight four here. First, we identified the Mis12 centromere complex components DSN1 and PMF1 in the BUB1 and MAD1L1 purifications (Figures 4A, S7A, and S10A). The Mis12 complex is comprised of PMF1, MIS12, DSN1, and NSL1 (42-44) and genetic and biochemical studies have shown that it coordinates communication from the outer kinetochore to the centromeric DNA in the inner kinetochore (44-46). Unexpectedly, PMF1 was also identified in the BUB3 purification (Figures 4A and S8A). To our knowledge there have been no previous reports of a direct association between BUB3 and the Mis12 complex. Therefore, this BUB3-PMF1 association could indicate a novel direct interaction or simply that these proteins reside within close proximity at the kinetochore. Of interest, the Mis12 complex recruits KNL1 to the kinetochore, which functions as a scaffold for the recruitment of BUB3 that subsequently recruits additional SAC components (4,38,47). Consistently, we observed the association of KNL1 with BUB1, BUB3, BUBR1, and MAD1L1 (Figure 4A). These associations were previously reported, as summarized in the Figure 4B BioGRID network, and had been established to have a role in checkpoint

activation (41,48-50) (reviewed in (5)). Additionally, MAD2L1 was not found to associate with KNL1 and to our knowledge a KNL1-MAD2L1 interaction has not been reported.

Second, minor components of the Astrin-Kinastrin complex (PLK1, DYNLL1, and SGO2) (51) were found to associate with all of the core SAC proteins (Figures 4A, S7A, S8A, S9A, S10A, and S11A). The Astrin-Kinastrin complex is important for aligning and attaching microtubules to kinetochores (51-53). Previous studies showed that depletion of BUB1 led to the delocalization of PLK1 and SGO2 from the kinetochores during prometaphase (54,55). Additionally, the BUB1 kinase activity was shown to be important for SGO2 kinetochore localization (56) and for the proper localization of BUB1 to the kinetochore (55) and pharmacological inhibition of the BUB1 kinase activity led to delocalization of SGO2 away from kinetochores (57). However, whether the BUB1 kinase activity was required for PLK1 kinetochore localization remained unknown. To address this, we first sought to confirm that PLK1 and SGO2 were mislocalized in BUB1-depleted cells. HeLa cells were treated with control siRNA (siControl) or BUB1-targeting siRNA (siBUB1) capable of depleting BUB1 protein levels (Figure 5A). Immunofluorescence microscopy of these cells showed that BUB1 was absent from kinetochores in siBUB1-treated cells (Figure 5B). Additionally, the siBUB1 treatment reduced the levels of kinetochore-localized PLK1 and SGO2 (Figure 5C,D). Next, we asked if the BUB1 kinase activity was required for PLK1 and SGO2 kinetochore localization. RPE cells were treated with control DMSO vehicle or the recently developed BUB1 kinase selective inhibitor BAY 1816032 (21) and the localization of PLK1 and SGO2 was assessed in mitotic cells. In comparison to the control DMSO treatment, treatment with BAY 1816032 led to a reduction in the levels of kinetochore-localized PLK1 and SGO2 (Figure 5E,F).

Additionally, treatment of BioID2-BUB1 expressing HeLa cells with BAY 1816032 also led to a reduction in the levels of kinetochore-localized BioID2-BUB1 (Figure 5G). This data indicated that the BUB1 kinase activity was important for its proper localization to kinetochores and for the localization of the Astrin-Kinastrin minor complex components PLK1 and SGO2 to the kinetochore.

Third, we identified CENPV as a MAD2L1 associating protein (Figure 4A). CENPV was identified in a proteomic screen for novel components of mitotic chromosomes (58) and was later shown to localize to kinetochores early in mitosis and to have a major role in directing the chromosomal passenger complex (CPC) subunits Aurora B and INCENP to the kinetochore (50,59). Although BUB1 has been shown to be important for the recruitment of the CPC to kinetochores (60), we are unaware of any reports of MAD2L1 being involved in this process. Interestingly, MAD2L1 has been shown to regulate the relocation of the CPC from centromeres through its inhibition of MKLP2, which is essential for proper cytokinesis (61). Thus, it is possible MAD2L1 could also be regulating CPC localization to kinetochores through its association with CENPV.

Fourth, components of the nuclear pore complex were found to associate with MAD1L1 and MAD2L1 (Figure S5). To better visualize these nuclear pore associated proteins, we performed a proximity protein mapping analysis for each of the core SAC proteins using the nuclear pore related GO annotations (see Table S8 for a list of nuclear pore related GO IDs) (Figure S16). This analysis revealed that MAD1L1 had associations with nuclear pore basket components including TPR, NUP153, NUP50, and other components of the nuclear pore that are in close proximity to the nuclear basket like ELYS/AHCTF1 (also known as MEL-28 in *C. elegans*) and NUP107 (Figure S16A). These

data support previous studies in humans and other organisms that have shown that MAD1L1 associates with TPR, NUP153, ELYS, and NUP107 and is important for generating the MAD1L1-MAD2L1 complex in early mitosis to establish the SAC (62-68). Similarly, MAD2L1 was found to associate with TPR (previously verified in (63)), NUP50, Nup153, NUP210 and ELYS (Figure S16A). Of interest, we did not detect associations between other core SAC proteins (BUB1; BUB3; BUBR1) and nuclear pore basket proteins. These data are consistent with a model where MAD1L1 makes multiple direct contacts with the nuclear pore basket complex subunits and MAD2L1 is in close proximity to NUP153 and NUP50 due to its binding to MAD1L1. We note that ELYS was found in both the MAD1L1 and MAD2L1 proximity maps (Figure S16A). ELYS was discovered in a proteomic screen for NUP107-160 complex binding partners and was shown to localize to nuclear pores in the nuclear lamina during interphase and to kinetochores during early mitosis, similar to the NUP107-160 complex (69). More recently, ELYS was shown to function as a scaffold for the recruitment of Protein Phosphatase 1 (PP1) to the kinetochore during M-phase exit, which was required for proper cell division (70,71). Due to ELYS's roles at the kinetochore and an identified yeast two-hybrid interaction between *C. elegans* MEL-28 (ELYS in humans) and MDF-1 (MAD1L1 in humans) (65), we sought to determine if MAD1L1 and MAD2L1 were binding directly to ELYS. First, we performed MYC immunoprecipitations from mitotic protein extracts prepared from BioID2, BioID2-MAD1L1, and BioID2-MAD2L1 expressing cell lines that had been arrested in mitosis. Indeed, ELYS immunoprecipitated with both BioID2-MAD1L1 and BioID2-MAD2L1, albeit weakly, in these mitotic extracts (Figure 6A). Next, we sought to assess these interactions in a cell-free *in vitro* expression system. Although a validated full-length ELYS cDNA

vector was not available and could not be generated, we were able to generate a MYC-tagged ELYS N-terminal fragment vector that expressed the first 46 amino acids of ELYS. This ELYS N-terminal fragment bound to FLAG-MAD1L1, but not FLAG-MAD2L1 (Figure 6B). Together, these data indicated that ELYS associated with MAD1L1 and MAD2L1 in mitotic cell extracts and that MAD1L1 bound to the ELYS N-terminal fragment *in vitro*.

Core SAC Proteins in Cellular Homeostasis

It's important to note that most of the core SAC proteins have been shown to have roles in cellular homeostasis independent of their role in the SAC, which are predominantly mediated through protein-protein interactions with non-kinetochore proteins. Many of these associations were present in the individual core SAC protein proximity maps where GO annotations were not applied (Figure S5). Consistently, Reactome pathway analysis of the core SAC protein proximity protein network showed that many of the SAC associated proteins had roles in numerous pathways important for cellular homeostasis including the cell cycle, DNA repair, and gene expression (Figure 4C). We encourage researchers interested in non-mitotic SAC protein functions to explore the SAC protein proximity association networks to gain further insights into these pathways.

CONCLUSIONS

The SAC is an important signaling pathway that is critical for proper cell division, which functions with great precision in a highly orchestrated manner (2). Due to the dynamic nature of the associations between core SAC proteins and the complexes and

subcomplexes that they form, it has been difficult to generate a proteomic network view of the proteins that are in close proximity and that interact with core SAC proteins. Here, we have established an inducible BiID2-tagging system that allowed for the transient expression of BiID2-tagged core SAC proteins (BUB1, BUB3, BUBR1, MAD1L1, and MAD2L1), which bypasses issues associated with long-term overexpression of key cell division proteins that can compromise cellular homeostasis. We coupled this system to a proximity labeling proteomic approach to systematically define a proximity protein association map for each of the core SAC proteins. These proximity maps were integrated to generate a core SAC protein proximity protein network. The coupling of the proximity maps/network with curated functional databases like CORUM, GeneOntology, BioGRID, and Reactome allowed for a systems level bioinformatic analysis of the associations within these maps/network. To our knowledge this is the first systematic characterization of the core SAC proteins by proximity-based proteomics.

Our analysis recapitulated many of the core SAC protein-protein interactions, sub-complexes, and complexes that had been previously described. Importantly, it also identified numerous novel associations that warrant further examination. Among these is ELYS, which associated with MAD1L1 and MAD2L1. Although an interpretation of these associations could be that MAD1L1 and MAD2L1 associate with ELYS at the nuclear pore in preparation for mitotic entry and SAC activation, we favor a model where ELYS may be important for the recruitment of SAC proteins to the kinetochore and/or for checkpoint activation. Future studies aimed at addressing these models should bring clarity to the potential role of ELYS in SAC functioning and cell division. Of interest, previous studies had shown the importance of BUB1 for the localization of the Astrin-Kinastrin minor

complex proteins to the kinetochore (51-54) and our analysis further determined that the BUB1 kinase activity was important for this function. Together, these data indicate that BUB1 may have a central organizing role not only in SAC activation and function, but in SAC silencing and mediating the transition from metaphase to anaphase through its association with the Astrin-Kinastrin minor complex (Figure 5H).

We note that there are limitations to the BioID2 approach (for review see (72)). Although our analysis was conducted from mitotic-enriched populations of cells to enrich for mitotic protein associations, the biotinylation process is conducted over the time-frame of hours and some identified associations could represent associations that take place outside of mitosis. These associations could inform on the non-mitotic functions of core SAC proteins, which is a rapidly growing field. Moving forward recent developments in BioID2 technology such as the mini-turboID system should help to resolve proximity associations in a time-dependent manner, as labeling occurs within minutes (73). Our analysis also employed N-terminal BioID2-tagging and a similar approach using C-terminal tagging of core SAC proteins could lead to different results. Additionally, it is important to note that BioID systems do not identify all known interactions of any specific bait protein. For example, we did not identify the MAD1L1-MAD2L1 interaction in our BioID2 analysis, which is consistent with a previous BioID analysis of MAD1L1 (41). Interestingly, we were able to detect the MAD1L1-MAD2L1 interaction when we performed immunoprecipitations with BioID2-MAD1L1 and BioID2-MAD2L1 and immunoblotted for endogenous MAD1L1 or MAD2L1 (Figures S4B and 6A). This indicates that BioID2-MAD1L1 is capable of binding to MAD2L1, but is not able to biotinylate it efficiently. There are many reasons why similar phenomena may occur with

other protein pairs and these include a low abundance of surface exposed lysines on prey proteins (whether absent from the protein surface or buried within a protein complex) and the orientation of the protein interaction could preclude access to lysines on prey proteins (72). Nonetheless, BioID systems have been invaluable to understanding cellular process and the architecture of cellular structures (14,19,74-76).

To facilitate the use and interrogation of the core SAC protein proximity maps/network generated in this study, all mass spectrometry data and R scripts used to analyze the data have been deposited in open access databases that are freely available to the scientific community (see Experimental Procedures). These tools will enable researches to define novel associations and to generate testable hypotheses to further advance the current understanding of SAC protein function and regulation.

SUPPORTING INFORMATION

The Supporting Information is available free of charge at: (link provided by ACS)

Workflow for generating BioID2 vectors and cell lines; workflow of mass spectrometry data acquisition and analysis; supporting data characterizing cell lines and biochemical purification controls; combined Cytoscape protein association maps of selected SAC proteins with no GO terms applied; Cytoscape protein association maps for each SAC protein with applied GO terms; kinetochore protein enrichment analysis; uncropped immunoblots for all figures (PDF)

List of reagents used (XLSX)

List of primers used (XLSX)

List of vectors generated (XLSX)

Summary of all identified peptides from all BioID2 purifications (XLSX)

Summary of all identified proteins from all BioID2 purifications (XLSX)

Summary of peptides for all proteins that were identified with one peptide sequence (XLSX)

Summary of significant SAC protein proximity associated proteins (XLSX)

List of Gene Ontology (GO) annotations used in the core SAC protein proximity association network analyses (XLSX)

NOTES

The authors declare no competing financial interest. The pGBioID2-27 and pGBioID2-47 vectors have been deposited at Addgene (Addgene IDs: 140276 and 140277 respectively) and are available to the scientific community. All raw mass spectrometry files can be accessed at the UCSD Center for Computational Mass Spectrometry MassIVE datasets <ftp://MSV000084975@massive.ucsd.edu>. All R scripts used in this study are freely available at GitHub <https://github.com/uclatorreslab/MassSpecAnalysis>.

ACKNOWLEDGEMENTS

This material is based upon work supported by the National Institutes of Health NIGMS grant numbers R35GM139539 and R01GM117475 to J.Z.T., any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Institutes of Health NIGMS. Y.A.G. and K.M.C. were supported by the UCLA Tumor Cell Biology Training Program (USHHS Ruth L. Kirschstein Institutional National Research Service Award # T32CA009056). This work was supported in part by a grant to The University of California, Los Angeles from the Howard Hughes Medical Institute through the James H. Gilliam Fellowships for Advanced Study Program (E.F.V), by a UCLA Molecular Biology Institute Whitcome Fellowship (E.F.V.) and a NIH P30DK063491 grant (J.P.W.).

FIGURES

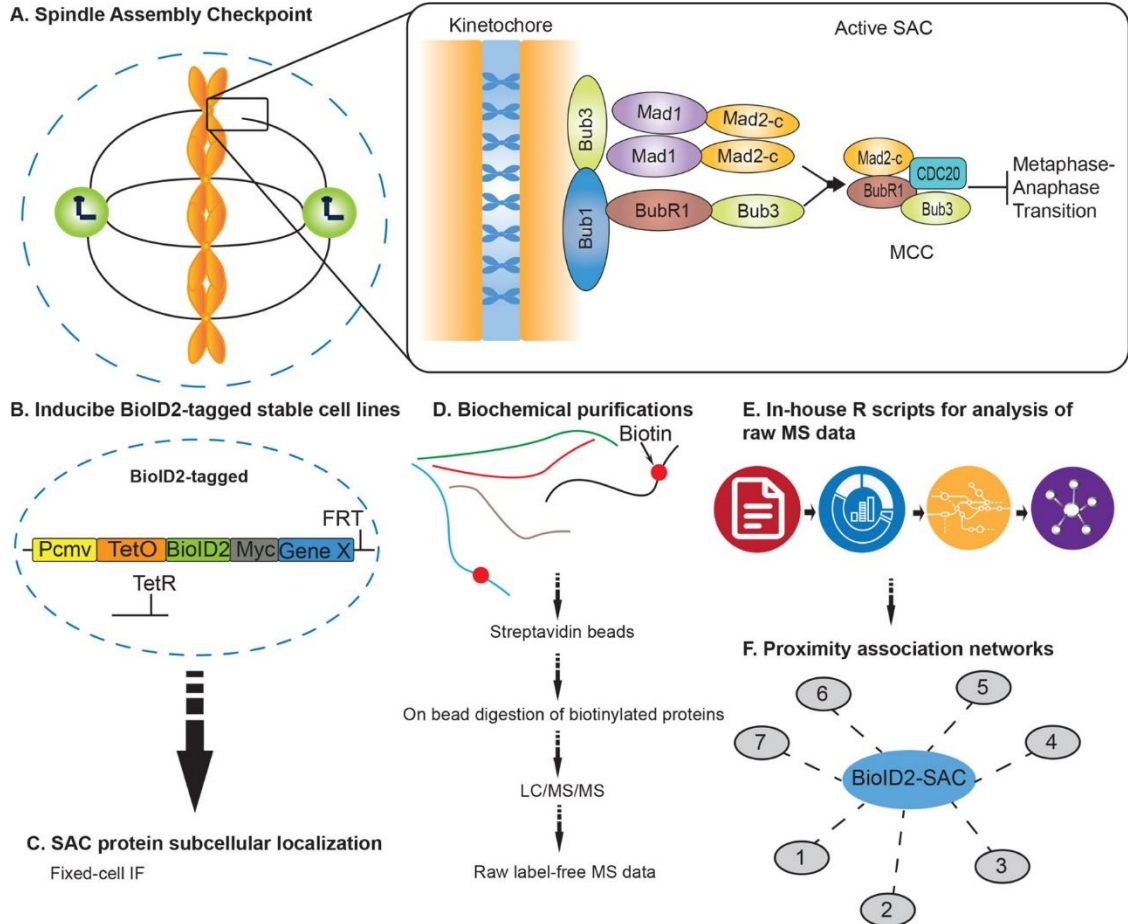


Figure 1. Overview of the approach to generate core SAC protein BiolD2 proximity association networks. (A) Schematic of the core spindle assembly checkpoint (SAC) components BUB1, BUB3, BUBR1, MAD1L1, and MAD2L1 that localize to the kinetochore region during early mitosis. MCC denotes mitotic checkpoint complex. (B) Generation of inducible BiolD2-tagged stable cell lines for each core SAC protein. (C) Fixed-cell immunofluorescence microscopy to analyze BiolD2-tagged SAC protein subcellular localization in time and space. (D) Biochemical purifications; affinity purification of biotinylated proteins and identification of proteins by LC/MS/MS. (E)

Computational analysis of raw mass spectrometry data using in-house R scripts. (F)
 Generation of high-confidence SAC protein proximity association networks.

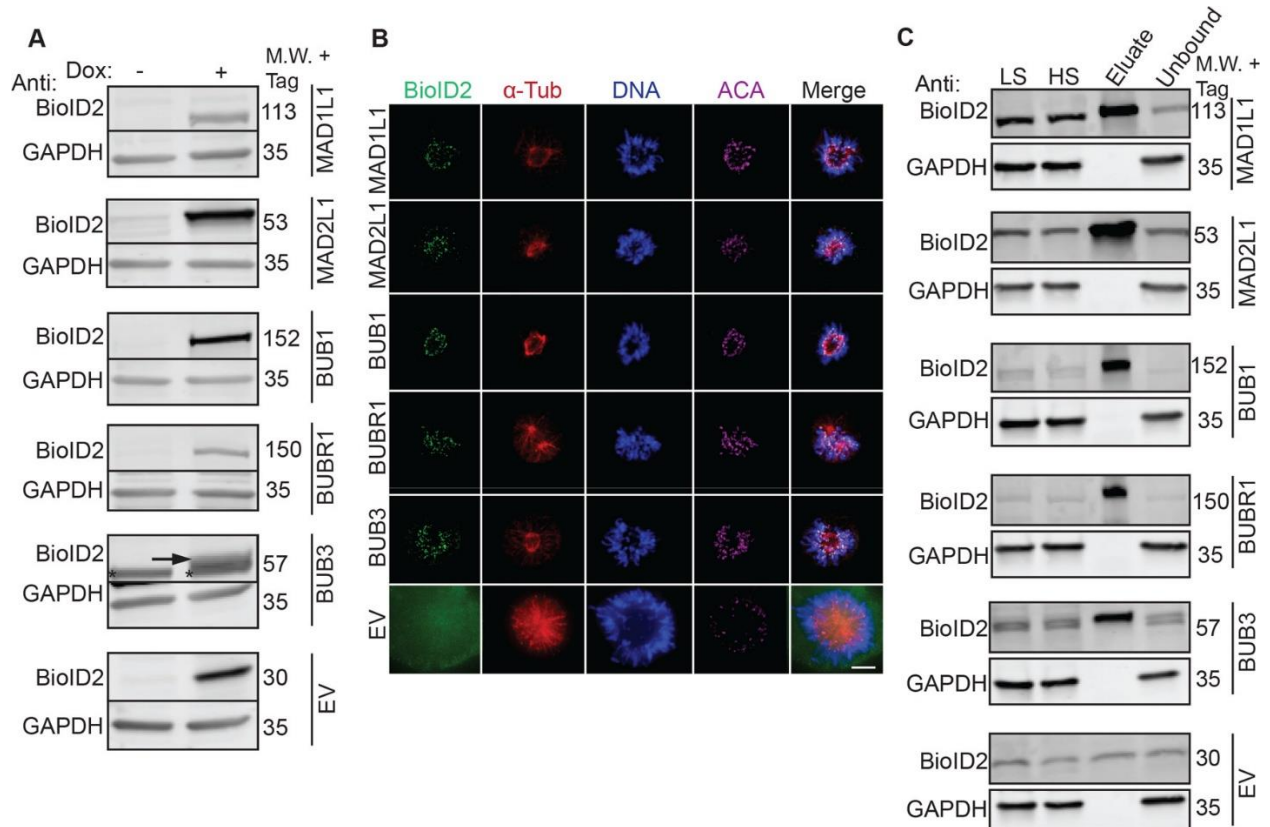


Figure 2. Establishment of inducible BioID2-tagged SAC protein (BUB1, BUB3, BUBR1, MAD1L1 and MAD2L1) stable cell lines and biochemical purifications. (A) Immunoblot analysis of extracts from doxycycline (Dox)-inducible BioID2-tag alone (EV, empty vector) or BioID2-tagged SAC protein (BUB1; BUB3; BUBR1; MAD1L1; MAD2L1) expression cell lines in the absence (-) or presence (+) of Dox for 16 hours. For each cell line, blots were probed with anti-BioID2 (to visualize the indicated BioID2-tagged SAC protein) and anti-GAPDH as a loading control. M.W. indicates molecular weight. Note that BioID2-tagged SAC proteins are only expressed in the presence of Dox. The

arrow points to the induced BioID2-BUB3 protein band and the asterisk denotes a non-specific band recognized by the anti-BioID2 antibody. (B) Fixed-cell immunofluorescence microscopy of the BioID2-tag alone (EV) or the indicated BioID2-tagged SAC proteins during prometaphase, a time when the SAC is active. HeLa BioID2-tagged protein expression cell lines were induced with Dox for 16 hours, fixed and stained with Hoechst 33342 DNA dye and anti-BioID2, anti- α -Tubulin and anti-centromere antibodies (ACA). Bar indicates 5 μ m. Note that all BioID2-tagged SAC proteins localize to the kinetochore region (overlapping with the ACA signal), whereas the BioID2-tag alone (EV) was absent from kinetochores. (C) Immunoblot analysis of BioID2 biochemical purifications from cells expressing the indicated BioID2-tagged SAC proteins or the BioID2-tag alone (EV). For each cell line, blots were probed with anti-BioID2 (to visualize the indicated BioID2-tagged SAC protein) and anti-GAPDH as a loading control. M.W. indicates molecular weight. LS indicates low speed supernatant, HS indicates high speed supernatant. Uncropped immunoblots are provided in Figures S17 and S18.

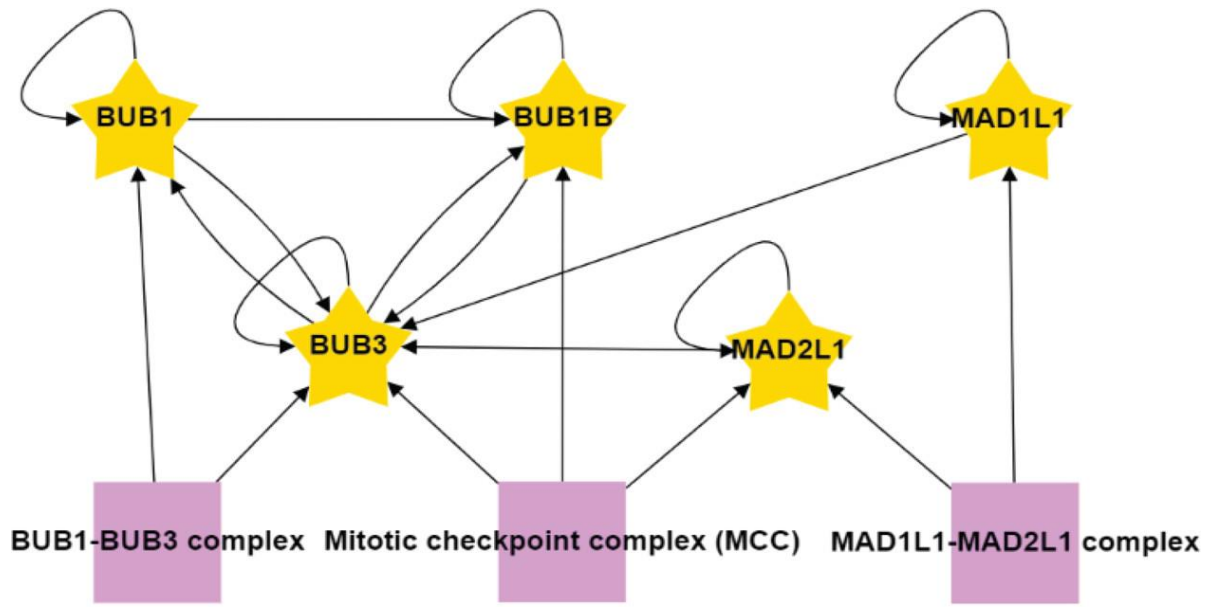


Figure 3. Associations among the core SAC proteins identified in the proximity protein network. The associations between each of the core SAC proteins (BUB1; BUB3; BUBR1; MAD1L1; MAD2L1) were isolated from the unified core SAC protein proximity association network (Figure S6). Purple boxes highlight protein complexes known to assemble with core SAC proteins as annotated by the CORUM database. Arrows indicate the direction of the detected associations.

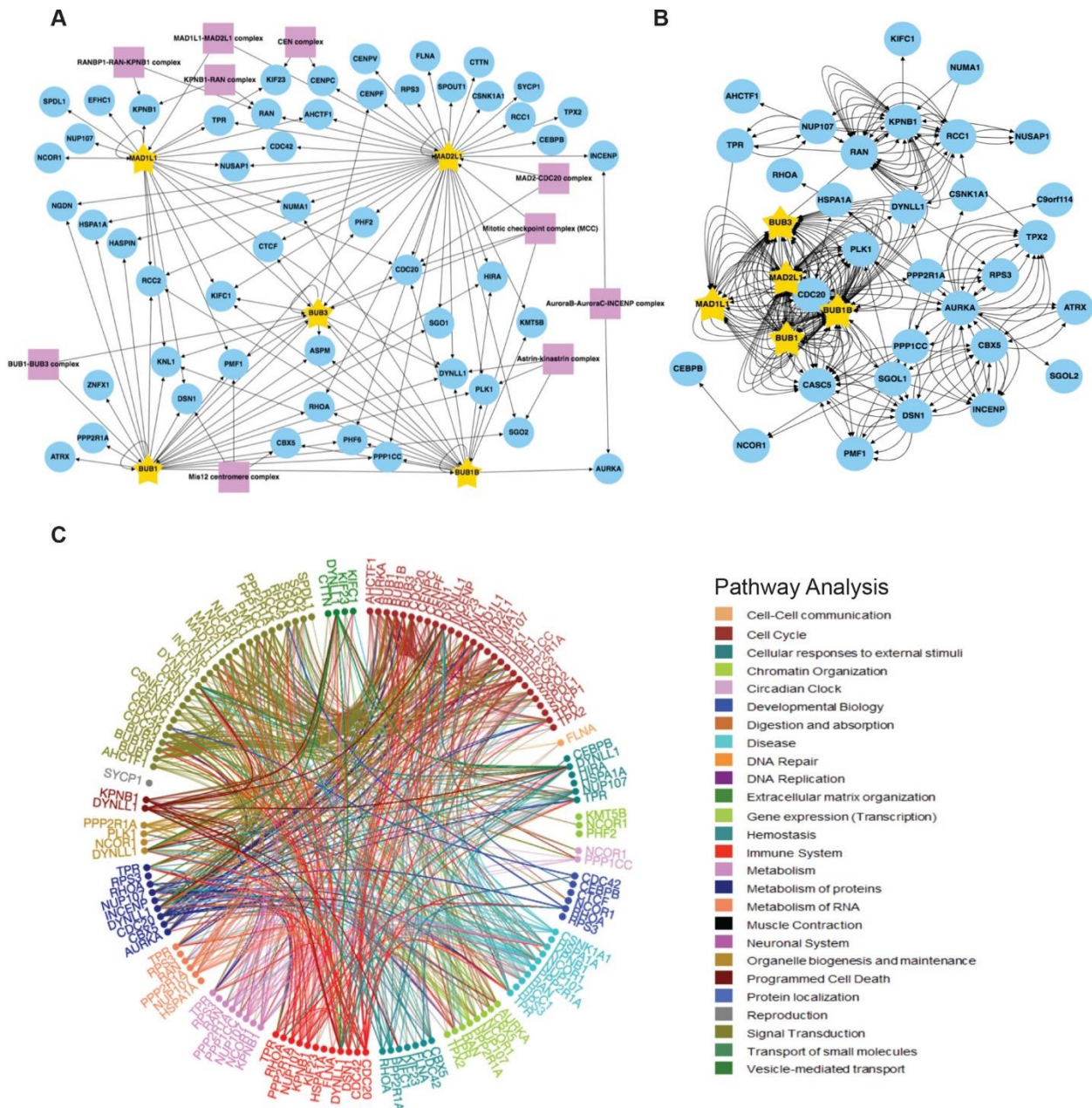


Figure 4. SAC protein BiOId2 kinetochore/mitotic spindle assembly/ centromere proximity association network. (A) Individual core SAC protein (BUB1; BUB3; BUBR1; MAD1L1; MAD2L1) proximity protein maps were compiled and subjected to kinetochore, mitotic spindle assembly, and centromere GO annotation analysis along with a COURM complex annotation analysis to generate a core SAC protein kinetochore/mitotic spindle assembly/ centromere proximity association network. Purple

boxes highlight kinetochore, mitotic spindle assembly, and centromere associated protein complexes present in the network. Arrows indicate the direction of the detected interactions. For a list of GO terms used see Table S8. (B) The core SAC protein kinetochore/mitotic spindle assembly/ centromere proximity association network was analyzed with BioGRID to reveal previously verified protein associations. Each arrow indicates an experimentally annotated interaction curated in the BioGRID database. Direction of arrows indicate an annotated interaction from a bait protein to the prey. (C) Reactome pathway analysis of the core SAC protein kinetochore/mitotic spindle assembly/ centromere proximity association network. The Reactome circular interaction plot depicts the associations between the identified proteins within the SAC protein kinetochore/mitotic spindle assembly/centromere proximity association network and the corresponding pathways in which they function. Legend presents the color-coded pathways that correspond to the circular interaction plots.

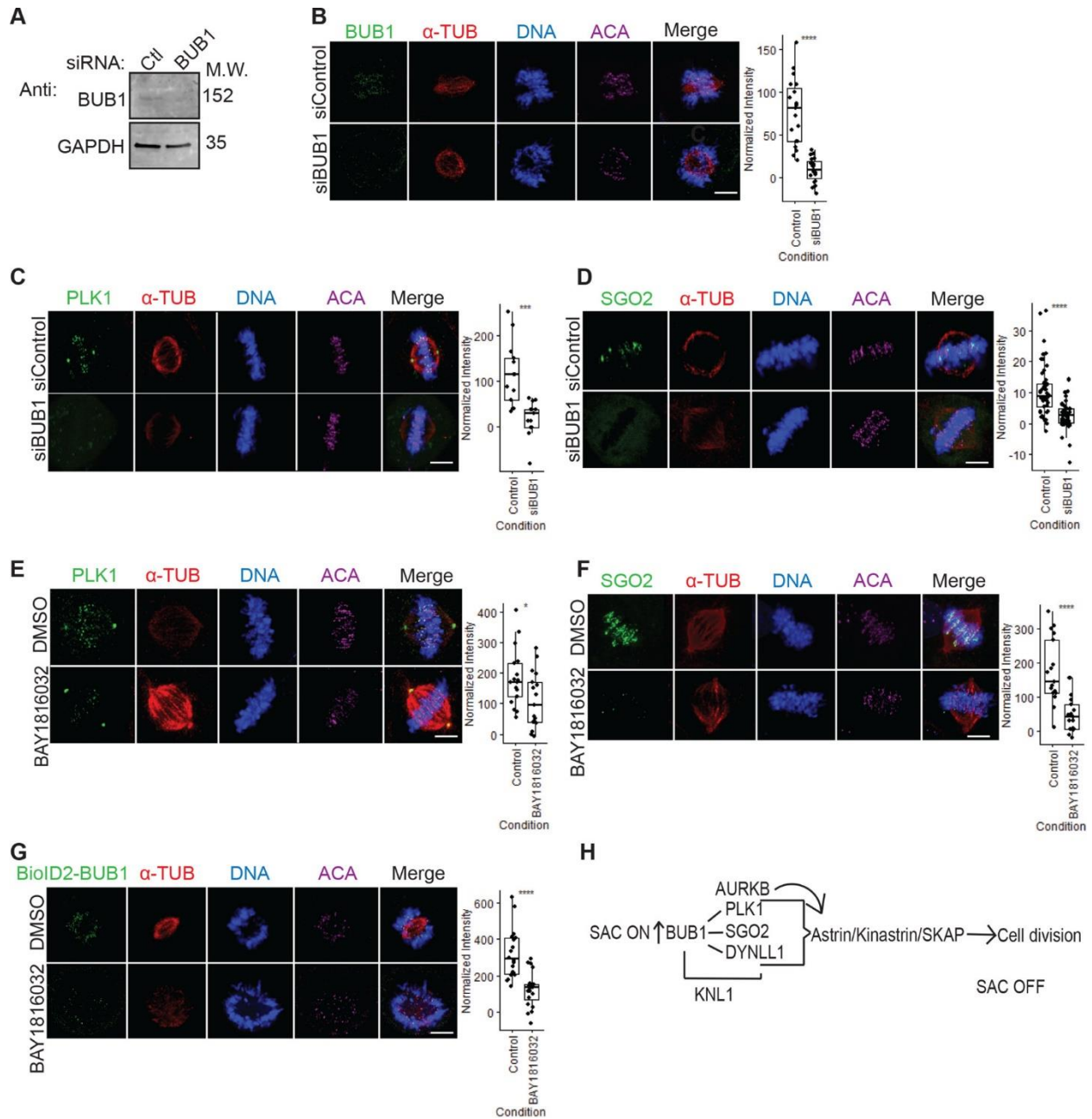


Figure 5. BUB1 as a hub for organizing the metaphase to anaphase transition. (A) Immunoblot analysis of protein extracts isolated from HeLa cells treated with control (Ctl) or BUB1 siRNA. GAPDH was used as a loading control. (B-D) Fixed-cell immunofluorescence microscopy of mitotic HeLa cells treated with control siRNA (siControl) or siRNA targeting BUB1 (siBUB1). Cells were fixed and stained with

Hoechst 33342 DNA dye and anti-BUB1 (B), anti-PLK1 (C), or anti-SGO2 (D) antibodies, along with anti- α -Tubulin and anti-centromere antibodies (ACA). Bars indicate 5 μ m. Box plots on the right of each panel show the quantification of the normalized fluorescence intensity for kinetochore-localized BUB1 (B), PLK1 (C), or SGO2 (D) and **** denotes a P-value < 0.001. (E-F) Same as in A, except that RPE cells were used and treated with control DMSO vehicle or the BUB1 kinase inhibitor BAY 1816032. Note that the levels of kinetochore-localized PLK1 (E) and SGO2 (F) decrease in BAY 1816032-treated cells. Bars indicate 5 μ m. Box plots on the right of each panel show the quantification of the normalized fluorescence intensity for kinetochore-localized PLK1 (E, * indicates P-value of 0.027) or SGO2 (F, **** indicates P-value < 0.001). (G) Same as in E-F, except that a HeLa BioID2-BUB1 expressing cell line was used. Bar indicates 5 μ m. Box plot shows the quantification of the normalized fluorescence intensity for kinetochore-localized BioID2-BUB1, **** indicates P-value < 0.001. (H) Model of BUB1 as an organizer of the metaphase to anaphase transition. BUB1 is critical for SAC protein binding to KNL1 to establish the SAC response and is also critical for the recruitment of the Astrin-Kinastrin minor complex, which is essential for the metaphase to anaphase transition.

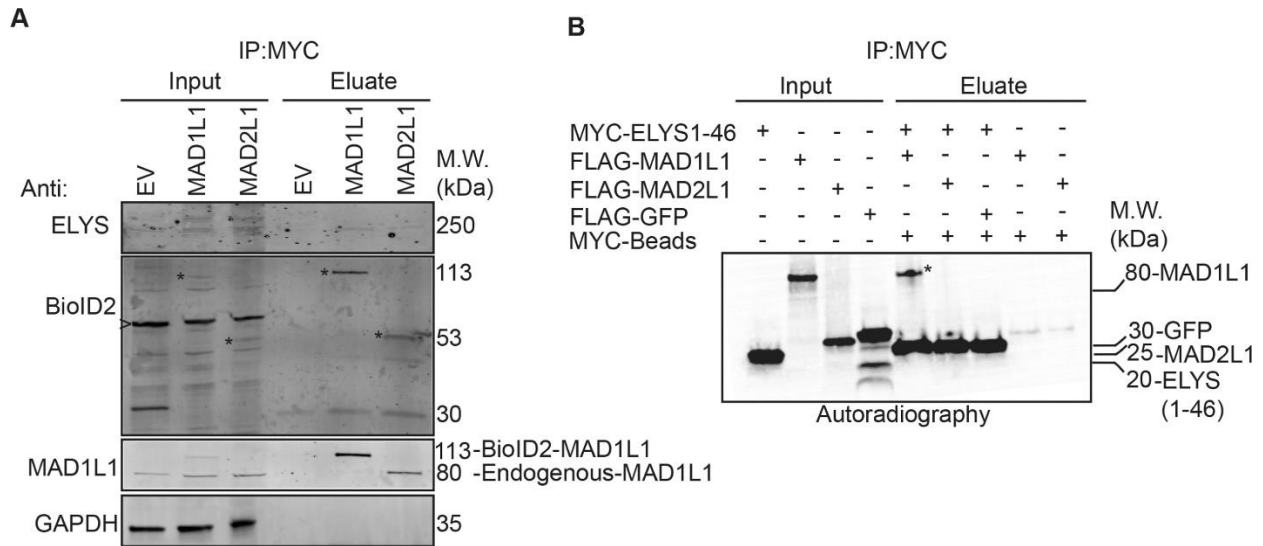


Figure 6. ELYS binds to MAD1L1 and MAD2L1 in mitotic cell lysates and to MAD1L1 *in vitro*. (A) BioID2-Myc (empty vector, EV), BioID2-Myc-MAD1L1, or BioID2-Myc-MAD2L1 inducible HeLa stable cell lines were induced with Dox and treated with 100 nM Taxol to arrest cells in mitosis. Mitotic cell lysates were then used for Myc immunoprecipitations and subjected to immunoblot analysis with the indicated antibodies. Note that endogenous ELYS immunoprecipitates with BioID2-Myc-tagged MAD1L1 and MAD2L1. Asterisks indicate BioID2-Myc-MAD1L1 or BioID2-Myc-MAD2L1 in the inputs or eluates. Arrow head indicates non-specific background band recognized by the anti-BioID2 antibody. (B) ³⁵S-radiolabeled Myc-ELYS N-terminal fragment (ELYS1-46, first 46 amino acids), FLAG-MAD1L1, FLAG-MAD2L1, and FLAG-GFP (control) were used in *in vitro* binding assays. Myc immunoprecipitations were resolved by western blotting and the blots were analyzed by autoradiography. Note that the ELYS N-terminal fragment binds to MAD1L1 and not MAD2L1.

REFERENCES

1. Holland, A. J., and Cleveland, D. W. (2009) Boveri revisited: chromosomal instability, aneuploidy and tumorigenesis. *Nat Rev Mol Cell Biol* **10**, 478-487
2. Musacchio, A., and Salmon, E. D. (2007) The spindle-assembly checkpoint in space and time. *Nat Rev Mol Cell Biol* **8**, 379-393
3. Gelens, L., Qian, J., Bollen, M., and Saurin, A. T. (2018) The Importance of Kinase-Phosphatase Integration: Lessons from Mitosis. *Trends Cell Biol* **28**, 6-21
4. Foley, E. A., and Kapoor, T. M. (2013) Microtubule attachment and spindle assembly checkpoint signalling at the kinetochore. *Nat Rev Mol Cell Biol* **14**, 25-37
5. Liu, S. T., and Zhang, H. (2016) The mitotic checkpoint complex (MCC): looking back and forth after 15 years. *AIMS Mol Sci* **3**, 597-634
6. Sudakin, V., Chan, G. K., and Yen, T. J. (2001) Checkpoint inhibition of the APC/C in HeLa cells is mediated by a complex of BUBR1, BUB3, CDC20, and MAD2. *J Cell Biol* **154**, 925-936
7. Stemmann, O., Zou, H., Gerber, S. A., Gygi, S. P., and Kirschner, M. W. (2001) Dual inhibition of sister chromatid separation at metaphase. *Cell* **107**, 715-726
8. Matson, D. R., and Stukenberg, P. T. (2011) Spindle poisons and cell fate: a tale of two pathways. *Mol Interv* **11**, 141-150
9. Ricke, R. M., and van Deursen, J. M. (2011) Correction of microtubule-kinetochore attachment errors: mechanisms and role in tumor suppression. *Semin Cell Dev Biol* **22**, 559-565

10. Hauf, S. (2013) The spindle assembly checkpoint: progress and persistent puzzles. *Biochem Soc Trans* **41**, 1755-1760
11. Corbett, K. D. (2017) Molecular Mechanisms of Spindle Assembly Checkpoint Activation and Silencing. *Prog Mol Subcell Biol* **56**, 429-455
12. Musacchio, A. (2015) The Molecular Biology of Spindle Assembly Checkpoint Signaling Dynamics. *Curr Biol* **25**, R1002-1018
13. Dou, Z., Prifti, D. K., Gui, P., Liu, X., Elowe, S., and Yao, X. (2019) Recent Progress on the Localization of the Spindle Assembly Checkpoint Machinery to Kinetochores. *Cells* **8**
14. Go, C. D., Knight, J. D. R., Rajasekharan, A., Rathod, B., Hesketh, G. G., Abe, K. T., Youn, J.-Y., Samavarchi-Tehrani, P., Zhang, H., Zhu, L. Y., Popiel, E., Lambert, J.-P., Coyaud, É., Cheung, S. W. T., Rajendran, D., Wong, C. J., Antonicka, H., Pelletier, L., Raught, B., Palazzo, A. F., Shoubridge, E. A., and Gingras, A.-C. (2019) A proximity biotinylation map of a human cell. *bioRxiv*, 796391
15. Firat-Karalar, E. N., and Stearns, T. (2015) Probing mammalian centrosome structure using BioID proximity-dependent biotinylation. *Methods Cell Biol* **129**, 153-170
16. Roux, K. J., Kim, D. I., and Burke, B. (2013) BioID: a screen for protein-protein interactions. *Curr Protoc Protein Sci* **74**, Unit 19 23
17. Mehus, A. A., Anderson, R. H., and Roux, K. J. (2016) BioID Identification of Lamin-Associated Proteins. *Methods Enzymol* **569**, 3-22

18. Kim, D. I., Jensen, S. C., Noble, K. A., Kc, B., Roux, K. H., Motamedchaboki, K., and Roux, K. J. (2016) An improved smaller biotin ligase for BioID proximity labeling. *Mol Biol Cell* **27**, 1188-1196
19. Gupta, G. D., Coyaud, E., Goncalves, J., Mojarad, B. A., Liu, Y., Wu, Q., Gheiratmand, L., Comartin, D., Tkach, J. M., Cheung, S. W., Bashkurov, M., Hasegan, M., Knight, J. D., Lin, Z. Y., Schueler, M., Hildebrandt, F., Moffat, J., Gingras, A. C., Raught, B., and Pelletier, L. (2015) A Dynamic Protein Interaction Landscape of the Human Centrosome-Cilium Interface. *Cell* **163**, 1484-1499
20. Senese, S., Cheung, K., Lo, Y. C., Gholkar, A. A., Xia, X., Wohlschlegel, J. A., and Torres, J. Z. (2015) A unique insertion in STARD9's motor domain regulates its stability. *Mol Biol Cell* **26**, 440-452
21. Siemeister, G., Mengel, A., Fernandez-Montalvan, A. E., Bone, W., Schroder, J., Zitzmann-Kolbe, S., Briem, H., Prechtel, S., Holton, S. J., Monning, U., von Ahsen, O., Johanssen, S., Cleve, A., Putter, V., Hitchcock, M., von Nussbaum, F., Brands, M., Ziegelbauer, K., and Mumberg, D. (2019) Inhibition of BUB1 Kinase by BAY 1816032 Sensitizes Tumor Cells toward Taxanes, ATR, and PARP Inhibitors In Vitro and In Vivo. *Clin Cancer Res* **25**, 1404-1414
22. Torres, J. Z., Miller, J. J., and Jackson, P. K. (2009) High-throughput generation of tagged stable cell lines for proteomic analysis. *Proteomics* **9**, 2888-2891
23. Bradley, M., Ramirez, I., Cheung, K., Gholkar, A. A., and Torres, J. Z. (2016) Inducible LAP-tagged Stable Cell Lines for Investigating Protein Function, Spatiotemporal Localization and Protein Interaction Networks. *J Vis Exp* **118**, 54870

24. Rappsilber, J., Mann, M., and Ishihama, Y. (2007) Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nat Protoc* **2**, 1896-1906
25. Cheung, K., Senese, S., Kuang, J., Bui, N., Ongpipattanakul, C., Gholkar, A., Cohn, W., Capri, J., Whitelegge, J. P., and Torres, J. Z. (2016) Proteomic Analysis of the Mammalian Katanin Family of Microtubule-severing Enzymes Defines Katanin p80 subunit B-like 1 (KATNBL1) as a Regulator of Mammalian Katanin Microtubule-severing. *Mol Cell Proteomics* **15**, 1658-1669
26. Ishihama, Y., Oda, Y., Tabata, T., Sato, T., Nagasu, T., Rappsilber, J., and Mann, M. (2005) Exponentially modified protein abundance index (emPAI) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein. *Mol Cell Proteomics* **4**, 1265-1272
27. Baez-Saldana, A., Zendejas-Ruiz, I., Revilla-Monsalve, C., Islas-Andrade, S., Cardenas, A., Rojas-Ochoa, A., Vilches, A., and Fernandez-Mejia, C. (2004) Effects of biotin on pyruvate carboxylase, acetyl-CoA carboxylase, propionyl-CoA carboxylase, and markers for glucose and lipid homeostasis in type 2 diabetic patients and nondiabetic subjects. *Am J Clin Nutr* **79**, 238-243
28. Giurgiu, M., Reinhard, J., Brauner, B., Dunger-Kaltenbach, I., Fobo, G., Frishman, G., Montrone, C., and Ruepp, A. (2019) CORUM: the comprehensive resource of mammalian protein complexes-2019. *Nucleic Acids Res* **47**, D559-D563

29. Stark, C., Breitkreutz, B. J., Reguly, T., Boucher, L., Breitkreutz, A., and Tyers, M. (2006) BioGRID: a general repository for interaction datasets. *Nucleic Acids Res* **34**, D535-539
30. Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., Harris, M. A., Hill, D. P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J. C., Richardson, J. E., Ringwald, M., Rubin, G. M., and Sherlock, G. (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* **25**, 25-29
31. Jassal, B., Matthews, L., Viteri, G., Gong, C., Lorente, P., Fabregat, A., Sidiropoulos, K., Cook, J., Gillespie, M., Haw, R., Loney, F., May, B., Milacic, M., Rothfels, K., Sevilla, C., Shamovsky, V., Shorser, S., Varusai, T., Weiser, J., Wu, G., Stein, L., Hermjakob, H., and D'Eustachio, P. (2020) The reactome pathway knowledgebase. *Nucleic Acids Res* **48**, D498-D503
32. Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* **13**, 2498-2504
33. Franz, M., Lopes, C. T., Huck, G., Dong, Y., Sumer, O., and Bader, G. D. (2016) Cytoscape.js: a graph theory library for visualisation and analysis. *Bioinformatics* **32**, 309-311
34. Xia, X., Gholkar, A., Senese, S., and Torres, J. Z. (2015) A LCMT1-PME-1 methylation equilibrium controls mitotic spindle size. *Cell Cycle* **14**, 1938-1947

35. Joglekar, A. P. (2016) A Cell Biological Perspective on Past, Present and Future Investigations of the Spindle Assembly Checkpoint. *Biology (Basel)* **5**
36. Zhang, G., Mendez, B. L., Sedgwick, G. G., and Nilsson, J. (2016) Two functionally distinct kinetochore pools of BubR1 ensure accurate chromosome segregation. *Nat Commun* **7**, 12256
37. Kulukian, A., Han, J. S., and Cleveland, D. W. (2009) Unattached kinetochores catalyze production of an anaphase inhibitor that requires a Mad2 template to prime Cdc20 for BubR1 binding. *Dev Cell* **16**, 105-117
38. Taylor, S. S., Ha, E., and McKeon, F. (1998) The human homologue of Bub3 is required for kinetochore localization of Bub1 and a Mad3/Bub1-related protein kinase. *J Cell Biol* **142**, 1-11
39. Overlack, K., Primorac, I., Vleugel, M., Krenn, V., Maffini, S., Hoffmann, I., Kops, G. J., and Musacchio, A. (2015) A molecular basis for the differential roles of Bub1 and BubR1 in the spindle assembly checkpoint. *Elife* **4**, e05269
40. De Antoni, A., Pearson, C. G., Cimini, D., Canman, J. C., Sala, V., Nezi, L., Mapelli, M., Sironi, L., Faretta, M., Salmon, E. D., and Musacchio, A. (2005) The Mad1/Mad2 complex as a template for Mad2 activation in the spindle assembly checkpoint. *Curr Biol* **15**, 214-225
41. Zhang, G., Kruse, T., Lopez-Mendez, B., Sylvestersen, K. B., Garvanska, D. H., Schopper, S., Nielsen, M. L., and Nilsson, J. (2017) Bub1 positions Mad1 close to KNL1 MELT repeats to promote checkpoint signalling. *Nat Commun* **8**, 15822
42. Maskell, D. P., Hu, X. W., and Singleton, M. R. (2010) Molecular architecture and assembly of the yeast kinetochore MIND complex. *J Cell Biol* **190**, 823-834

43. Petrovic, A., Pasqualato, S., Dube, P., Krenn, V., Santaguida, S., Cittaro, D., Monzani, S., Massimiliano, L., Keller, J., Tarricone, A., Maiolica, A., Stark, H., and Musacchio, A. (2010) The MIS12 complex is a protein interaction hub for outer kinetochore assembly. *Journal of Cell Biology* **190**, 835-852
44. Kline, S. L., Cheeseman, I. M., Hori, T., Fukagawa, T., and Desai, A. (2006) The human Mis12 complex is required for kinetochore assembly and proper chromosome segregation. *J Cell Biol* **173**, 9-17
45. Screpanti, E., De Antoni, A., Alushin, G. M., Petrovic, A., Melis, T., Nogales, E., and Musacchio, A. (2011) Direct Binding of Cenp-C to the Mis12 Complex Joins the Inner and Outer Kinetochore. *Curr Biol* **21**, 391-398
46. Obuse, C., Iwasaki, O., Kiyomitsu, T., Goshima, G., Toyoda, Y., and Yanagida, M. (2004) A conserved Mis12 centromere complex is linked to heterochromatic HP1 and outer kinetochore protein Zwint-1. *Nat Cell Biol* **6**, 1135-1141
47. Ghongane, P., Kapanidou, M., Asghar, A., Elowe, S., and Bolanos-Garcia, V. M. (2014) The dynamic protein Knl1 - a kinetochore rendezvous. *J Cell Sci* **127**, 3415-3423
48. Rodriguez-Rodriguez, J. A., Lewis, C., McKinley, K. L., Sikirzhyski, V., Corona, J., Maciejowski, J., Khodjakov, A., Cheeseman, I. M., and Jallepalli, P. V. (2018) Distinct Roles of RZZ and Bub1-KNL1 in Mitotic Checkpoint Signaling and Kinetochore Expansion. *Curr Biol* **28**, 3422-3429 e3425
49. Krenn, V., Wehenkel, A., Li, X., Santaguida, S., and Musacchio, A. (2012) Structural analysis reveals features of the spindle checkpoint kinase Bub1-kinetochore subunit Knl1 interaction. *J Cell Biol* **196**, 451-467

50. Mora-Santos, M. D., Hervas-Aguilar, A., Sewart, K., Lancaster, T. C., Meadows, J. C., and Millar, J. B. (2016) Bub3-Bub1 Binding to Spc7/KNL1 Toggles the Spindle Checkpoint Switch by Licensing the Interaction of Bub1 with Mad1-Mad2. *Curr Biol* **26**, 2642-2650
51. Dunsch, A. K., Linnane, E., Barr, F. A., and Gruneberg, U. (2011) The astrin-kinastrin/SKAP complex localizes to microtubule plus ends and facilitates chromosome alignment. *J Cell Biol* **192**, 959-968
52. Thein, K. H., Kleylein-Sohn, J., Nigg, E. A., and Gruneberg, U. (2007) Astrin is required for the maintenance of sister chromatid cohesion and centrosome integrity. *J Cell Biol* **178**, 345-354
53. Kern, D. M., Monda, J. K., Su, K. C., Wilson-Kubalek, E. M., and Cheeseman, I. M. (2017) Astrin-SKAP complex reconstitution reveals its kinetochore interaction with microtubule-bound Ndc80. *Elife* **6**
54. Ikeda, M., and Tanaka, K. (2017) Plk1 bound to Bub1 contributes to spindle assembly checkpoint activity during mitosis. *Sci Rep* **7**, 8794
55. Asghar, A., Lajeunesse, A., Dulla, K., Combes, G., Thebault, P., Nigg, E. A., and Elowe, S. (2015) Bub1 autophosphorylation feeds back to regulate kinetochore docking and promote localized substrate phosphorylation. *Nat Commun* **6**, 8364
56. El Yakoubi, W., Buffin, E., Cladiere, D., Gryaznova, Y., Berenguer, I., Touati, S. A., Gomez, R., Suja, J. A., van Deursen, J. M., and Wassmann, K. (2017) Mps1 kinase-dependent Sgo2 centromere localisation mediates cohesin protection in mouse oocyte meiosis I. *Nat Commun* **8**, 694

57. Baron, A. P., von Schubert, C., Cubizolles, F., Siemeister, G., Hitchcock, M., Mengel, A., Schroder, J., Fernandez-Montalvan, A., von Nussbaum, F., Mumberg, D., and Nigg, E. A. (2016) Probing the catalytic functions of Bub1 kinase using the small molecule inhibitors BAY-320 and BAY-524. *Elife* **5**
58. Gassmann, R., Henzing, A. J., and Earnshaw, W. C. (2005) Novel components of human mitotic chromosomes identified by proteomic analysis of the chromosome scaffold fraction. *Chromosoma* **113**, 385-397
59. Tadeu, A. M., Ribeiro, S., Johnston, J., Goldberg, I., Gerloff, D., and Earnshaw, W. C. (2008) CENP-V is required for centromere organization, chromosome alignment and cytokinesis. *EMBO J* **27**, 2510-2522
60. Boyarchuk, Y., Salic, A., Dasso, M., and Arnaoutov, A. (2007) Bub1 is essential for assembly of the functional inner centromere. *J Cell Biol* **176**, 919-928
61. Lee, S. H., McCormick, F., and Saya, H. (2010) Mad2 inhibits the mitotic kinesin MKlp2. *J Cell Biol* **191**, 1069-1077
62. Scott, R. J., Lusk, C. P., Dilworth, D. J., Aitchison, J. D., and Wozniak, R. W. (2005) Interactions between Mad1p and the nuclear transport machinery in the yeast *Saccharomyces cerevisiae*. *Mol Biol Cell* **16**, 4362-4374
63. Lee, S. H., Sterling, H., Burlingame, A., and McCormick, F. (2008) Tpr directly binds to Mad1 and Mad2 and is important for the Mad1-Mad2-mediated mitotic spindle checkpoint. *Genes Dev* **22**, 2926-2931
64. Lussi, Y. C., Shumaker, D. K., Shimi, T., and Fahrenkrog, B. (2010) The nucleoporin Nup153 affects spindle checkpoint activity due to an association with Mad1. *Nucleus* **1**, 71-84

65. Rodenas, E., Gonzalez-Aguilera, C., Ayuso, C., and Askjaer, P. (2012) Dissection of the NUP107 nuclear pore subcomplex reveals a novel interaction with spindle assembly checkpoint protein MAD1 in *Caenorhabditis elegans*. *Mol Biol Cell* **23**, 930-944
66. Rodriguez-Bravo, V., Maciejowski, J., Corona, J., Buch, H. K., Collin, P., Kanemaki, M. T., Shah, J. V., and Jallepalli, P. V. (2014) Nuclear pores protect genome integrity by assembling a premitotic and Mad1-dependent anaphase inhibitor. *Cell* **156**, 1017-1031
67. Gonzalez-Aguilera, C., and Askjaer, P. (2012) Dissecting the NUP107 complex: multiple components and even more functions. *Nucleus* **3**, 340-348
68. Mossaid, I., and Fahrenkrog, B. (2015) Complex Commingling: Nucleoporins and the Spindle Assembly Checkpoint. *Cells* **4**, 706-725
69. Rasala, B. A., Orjalo, A. V., Shen, Z., Briggs, S., and Forbes, D. J. (2006) ELYS is a dual nucleoporin/kinetochore protein required for nuclear pore assembly and proper cell division. *Proc Natl Acad Sci U S A* **103**, 17801-17806
70. Hattersley, N., Cheerambathur, D., Moyle, M., Stefanutti, M., Richardson, A., Lee, K. Y., Dumont, J., Oegema, K., and Desai, A. (2016) A Nucleoporin Docks Protein Phosphatase 1 to Direct Meiotic Chromosome Segregation and Nuclear Assembly. *Dev Cell* **38**, 463-477
71. Hattersley, N., and Desai, A. (2017) The nucleoporin MEL-28/ELYS: A PP1 scaffold during M-phase exit. *Cell Cycle* **16**, 489-490

72. Samavarchi-Tehrani, P., Samson, R., and Gingras, A. C. (2020) Proximity Dependent Biotinylation: Key Enzymes and Adaptation to Proteomics Approaches. *Mol Cell Proteomics* **19**, 757-773
73. Branon, T. C., Bosch, J. A., Sanchez, A. D., Udeshi, N. D., Svinkina, T., Carr, S. A., Feldman, J. L., Perrimon, N., and Ting, A. Y. (2018) Efficient proximity labeling in living cells and organisms with TurboID. *Nat Biotechnol* **36**, 880-887
74. Youn, J. Y., Dunham, W. H., Hong, S. J., Knight, J. D. R., Bashkurov, M., Chen, G. I., Bagci, H., Rathod, B., MacLeod, G., Eng, S. W. M., Angers, S., Morris, Q., Fabian, M., Cote, J. F., and Gingras, A. C. (2018) High-Density Proximity Mapping Reveals the Subcellular Organization of mRNA-Associated Granules and Bodies. *Mol Cell* **69**, 517-532 e511
75. Coyaud, E., Mis, M., Laurent, E. M., Dunham, W. H., Couzens, A. L., Robitaille, M., Gingras, A. C., Angers, S., and Raught, B. (2015) BioID-based Identification of Skp Cullin F-box (SCF) β -TrCP1/2 E3 Ligase Substrates. *Mol Cell Proteomics* **14**, 1781-1795
76. Remnant, L., Booth, D. G., Vargiu, G., Spanos, C., Kerr, A. R. W., and Earnshaw, W. C. (2019) In vitro BioID: mapping the CENP-A microenvironment with high temporal and spatial resolution. *Mol Biol Cell* **30**, 1314-1325

CHAPTER 4

Future Directions: Predicting Contaminants in PPI/A Data Using Supervised Machine Learning

INTRODUCTION

Predictive analysis via machine learning has become a popular technique in the biological sciences, fueled by improvements in computer systems, data accessibility, and analytical techniques. These advances have given rise to many breakthroughs in numerous scientific areas including molecular biology (Camacho *et al.*, 2018), image data analysis (Shen *et al.*, 2017) and clinical practice (Rajkomar *et al.*, 2019). Of particular interest are network-specifics about pathways or disease states, since protein-protein interactions and associations (PPI/A) contribute significantly to our understanding of function in biological systems. In this area scientists have used databases like BioGRID (Oughtred *et al.*, 2019), a database of gene interactions, protein-protein interactions and chemical interactions to study how different biomolecules relate to each other. Understanding these relationships allows us to have a better understanding of healthy states which can be applied to train a machine learning algorithm to understand healthy network characteristics of these states. Then providing the algorithm with patient data, machine learning can be used to predict between healthy and disease states, allowing scientists to identify key characteristics that distinguish disease states (Mall *et al.*, 2017). Similarly, machine learning techniques can be applied to understanding network inference, especially when considering phenotypic differences in expression (Fuente, 2010).

More recently, machine learning methods have been used to improve many aspects of the proteoinformatic pipeline in PPI/A MS based experiments. Mass spectrometry-based proteomics is a powerful technique that allows for the qualitative and quantitative analysis of proteins in complex mixtures. However, there are computational challenges in peptide

identification, quantification, quality assessment and biological interpretation that prevent more sensitive and accurate identification and quantification of proteins (Sagawa *et al.*, 2020; Meyer, 2021). In simple terms, machine learning methods approximate the value or classify an object based on the mathematical relationship between object (x) and output (y) (Badillo *et al.*, 2020). By learning from the relationship between outputs and objects machine learning methods are then able to apply this statistical relationship to new examples. This can be applied to many aspects of the proteoinformatic pipeline, for example predicting retention time of unknown proteins considering information from known proteins. Machine learning methods have been used to predict fragment ion intensity, retention time, protein and peptide identification (Table 1) (Meyer, 2021).

Table 1. Examples of machine learning methods used for fragment ion, retention time, and protein/peptide identification predictions.

Year	Name	Prediction Type	Citation
2005	PeptideART	Fragment Ion	(Arnold <i>et al.</i> , 2006)
2017	pDeep	Fragment Ion	(Zhou <i>et al.</i> , 2017)
2019	DeepMass	Fragment Ion and Retention Time	(Tiwary <i>et al.</i> , 2019)
2003	N/A	Retention Time	(Petritis <i>et al.</i> , 2003)
2017	DeepRT	Retention Time	(Ma <i>et al.</i> , 2017)
2012	Barista	Protein/Peptide Identification	(Spivak <i>et al.</i> , 2012)
2020	DIA-NN	Protein/Peptide Identification	(Demichev <i>et al.</i> , 2020)

Machine learning methods can be classified into two main categories: supervised and unsupervised methods (Figure 1). Supervised methods have labelled data and their goal is to predict a value (either numerical or categorical) given new data (Badillo *et al.*, 2020). Unsupervised methods have unlabeled data and their goal is to explore the structure of the data to find some sort of pattern or order (Badillo *et al.*, 2020). Figure 1 summarizes common machine learning methods and how they relate to each other. Once the right algorithm is chosen, machine learning projects often follow similar steps that lead to adapting the machine learning method to the data structure at hand. Often the steps involve obtaining data, preparing data, training/applying a model, testing the model and finally using the test to improve the algorithm itself (Badillo *et al.*, 2020).

Analysis of protein-protein interaction/association experiments using mass spectrometry often involves a quantitative or semi-quantitative/qualitative analysis workflow. The abundance of each protein in bait purifications is compared to abundance values of proteins in a control purification. However, depending on the mass spectrometer and preparation of samples before mass spectrometry, not all proteins identified in the bait purification are identified in the control purification and vice versa. This common problem falls under a common problem in mass spectrometry coined the missing data problem (Jin *et al.*, 2021). This problem is particularly prevalent in label-free quantitative proteomics, and it reduces the accuracy and completeness of the data. To overcome this issue, the field has adopted imputation methods that simulate missing values in order to complete the data (Jin *et al.*, 2021). However, accuracy of imputation value is dependent

on the selection of imputation method creating a wide array of results in the downstream pipeline. Alternatively, databases have been created to overcome the missing data problem in PPI/A MS-based experiments. The most popular database is the contaminant repository for affinity purification-mass spectrometry data (CRAPome) that houses label free quantitative values for common controls in PPI/A experiments (Mellacheruvu *et al.*, 2013). However, because of the nature of the experiment, not every protein identified in an experiment will be present in the controls of the CRAPome, and so there is a question of whether or not a protein should be considered a contaminant. To overcome this issue, we suggest a more general algorithm using supervised machine learning that aims to create a profile of proteins commonly found in a certain type of controls and aims to then applies this understanding to predicting whether a protein should be a contaminant or not.

This work describes a supervised machine learning method aiming to predict contaminants in PPI/A MS-based data. The pipeline involves four major steps: 1. Identifying a control of interest in the CRAPome (Mellacheruvu *et al.*, 2013) and data scrapping sequence information for proteins in the control. 2. Creating features to describe proteins in the control, we show that amphiphilicity is a main predictor to separate contaminants in PPI/A data. 3. Create a predictive model using supervised machine learning methods, we compare random forest and svm two popular machine learning methods. 4. Apply the model to test data. Using this pipeline we achieved 90% accuracy with a high sensitivity of 87%. We suggest this method is beneficial to labs that have established common methods of doing PPI/A experiments and would like to apply another quality control filter by which to start analyzing the data from these experiments.

Overall, this method can be applied to any PPI/A MS-based control experiment to create a general algorithm that predicts contaminants mostly based on amphiphilicity.

RESULTS

Unsupervised data exploration

To create a predictive model that identifies contaminants in PPI/A MS-based data, we sought to identify features that commonly describe protein sequences through proteochemometric modeling. We hypothesized that hydrophobic and hydrophilic pockets in proteins were not only important in determining protein-protein interactions but also could be used to identify common features in contaminants. Hence, we calculated the Amphiphilic Pseudo-Amino Acid Composition (APseAAC) (Chou, 2001; Chou, 2005; Xiao *et al.*, 2021) for protein sequences. In simple terms, APseAAC aims to calculate the hydrophobicity and hydrophilicity of protein sequences while considering local and global sequence order (Chou, 2001; Chou, 2005). We first identified 114 common contaminants in a BiOLD control purification in the CRAPome (Mellacheruvu *et al.*, 2013), then compared those to 1,000 randomly selected proteins from the UniProt database (Apweiler *et al.*, 2004) and around 80 to 90 protein interactions/associations for each SAC protein. Using the ProtR (Xiao *et al.*, 2021) computational suite for proteochemometric modeling in R we calculated the APseAAC from protein sequences in each group.

We decided to first explore the data in an unsupervised fashion to determine if there are any differences between protein groups. To do so we took the average APseAAC values

for all proteins in each group and compared them to each other. Figure 2A demonstrates hydrophobic and hydrophilic values across the first ten principal components. Figure 2B shows the difference across average values when the control and BUB1 interacting proteins are compared to contaminants. The difference heatmap (Figure 2B) clearly demonstrates that there are numerical differences between contaminants and the other two groups demonstrating that hydrophobicity and hydrophilicity can be used as features in predicting contaminants from interactors or associations in BioID purifications.

Supervised machine learning

Using amphiphilicity as a predictor, we applied supervised machine learning and measured for accuracy. Using interactors from 4 main SAC components (BUB1, BUB1B, MAD1, MAD2) we combined interactors with 1,000 random proteins from the UniProt database. We then compared these proteins to the most common contaminants in the HeLa Cell Line BioID in the CRAPome database. For each SAC component, we divided the dataset into a training (70% of the data) and test (30% of the data) set in order to create and test the model for accuracy. Given that there was an imbalance in classes (114 contaminants vs. 1000+ random and interaction proteins) we decided to apply Synthetic Minority Oversampling Technique (SMOTE), a technique used to balance uneven numerical classes by using Euclidian distances between groups to create new data points, in the training set (Blagus and Lusa, 2013). We then applied the trained random forest model to the test data and measured for accuracy. The results suggest the model is about ~90% accurate at predicting contaminants in test data. Furthermore, the accuracy is consistent across different baits (Figure 3).

CONCLUSION AND FUTURE DIRECTIONS

Mass spectrometry is a powerful tool for cell biologists, specially when understanding the proteome of cellular events and compartments. Affinity purification and proximity mass spectrometry-based assays have become extremely popular techniques that aim to elucidate the proteome around cellular events and proteins of interest. However, these techniques often result in large amounts of data which can be overwhelming for biologists to understand. To address these issues scientists have applied quantitative data and statistical analysis to better understand which proteins might be contaminants in interaction/proximity MS-based data. Although these techniques are powerful and accurate in labeled MS-based data, they are less accurate in unlabeled MS data. Furthermore, these types of experiments are prone to the missing data problem in which proteins might be contaminants but because of the limitations in MS-based techniques are not identified in control purifications. To address these issues, we have started to apply machine learning techniques to identify protein contaminants in interaction/proximity-based purifications. Our results suggest that amphiphilicity is a predictor of protein contaminants, and that applying supervised machine learning methods like random forest results in accurate (90% accurate) prediction of protein contaminants in interaction/proximity-based experiments. Furthermore, we tested this method with 4 different unique protein baits, resulting in similar results suggesting our predictions are accurate and reproducible across a variety of baits. In the future, we hope to extend this research to incorporate other types of protein-based numerical features and test unsupervised machine learning methods to perform quality control in interaction/proximity MS-based data.

METHODS

Preparing PPI/A Data

In order to create and test a supervised machine learning model that predicts contaminants in PPI/A data, we first created a training dataset from which we could train the model and monitor its performance. We decided to scrape data from the contaminant repository for affinity purifications (CRAPome) database (Mellacheruvu *et al.*, 2013). We decided to specifically focus on a BioID control purification performed in human Hela cells incubated with Biotin since this type of experiment mostly resembles the type of experiments done in our lab. More specifically, we chose experiments CC539, CC540, CC541, CC542, CC543, and CC544 and aggregated the scores of all proteins identified in the control purifications by calculating the mean spectral count value for each protein. We then sought to determine which proteins were most present in the purification, and so we calculated the distribution of the mean spectral count value. Using bootstrap method (Henderson, 2005), we calculated the distribution of the mean spectral count for proteins in the selected purifications and calculated a range for 95% confidence interval (CI) of the mean. We calculated the mean value to be 5.2 spectral count units (SCU) with the 95% CI within 3.5 and 7.8 SCUs. We reasoned that the most prevalent proteins would lie to the right of the 95% CI and labeled proteins with a spectral count higher than 7.8 SCUs as contaminants. We identified a total 114 proteins as common contaminants for this type of BioID purifications. We then sought to identify a separate class of proteins that would be representative of non-contaminants in order to train a supervised machine learning algorithm. To do so, we identified 1000 random proteins in the UniProt database (Apweiler *et al.*, 2004) and used that as our second class by which to train the algorithm

with. Finally, to identify probable interactors, it was important to include protein-protein interactions and associations for baits of interest in the training set of the supervised machine learning model. Using the IntAct database (Hermjakob *et al.*, 2004) we identified protein-protein interactions and associations for baits of interest in which to include in the training set.

Feature Extraction

We sought to identify features that commonly describe protein sequences through proteochemometric modeling. We reasoned that proteins commonly identified as contaminants must have common hydrophobic and hydrophilic regions. With this in mind, we focused on the amphiphilic pseudo-amino acid composition (APseAAC) (Chou, 2001) a feature that aims to calculate amino acid composition and the distribution of hydrophobic and hydrophilic amino acids while considering sequence order (Chou, 2005). This feature has been used in predictive modeling to predict enzyme subfamily classes (Chou, 2005), cofactors of oxidoreductases (Zhang and Fang, 2008) and drug-target interactions (Wang *et al.*, 2020). In simple terms it aims to create discrete numerical values from sequence order information by calculating the amphiphilicity of groups of amino acids. We calculated this feature for all the proteins identified in the previous section using the R package ProtR, which is a package developed to calculate common numerical features in protein sequences for the purposes of machine learning workflows. For a more detailed explanation of how this is calculated please refer to the original papers explaining APseAAC (Chou, 2001; Chou, 2005) and the ProtR package (Xiao *et al.*, 2021).

Building and Testing a Predictive Model

Once the data was cleaned and annotated, we sought to apply supervised machine learning to predict contaminants in PPI/A data. We first selected a training set which was used to create the algorithm and a test set by which to test the accuracy of the algorithm. The training set was composed of 70% of the data accounting for classes which was selected at random (70% of each class in the annotated data). The test set was composed of the remaining 30% of the data. In order to account for the imbalance between classes while training the algorithm, we took advantage of a technique meant to create balanced data in high dimensional data called Synthetic Minority Oversampling Technique (Blagus and Lusa, 2013). Briefly, the algorithm considers the data points of the minority class in multidimensional space and creates new data points by measuring near Euclidian distances between data points. The balanced data was then used to train a random forest algorithm in R. Finally, the trained algorithm was tested against the test data and measured for accuracy.

FIGURES

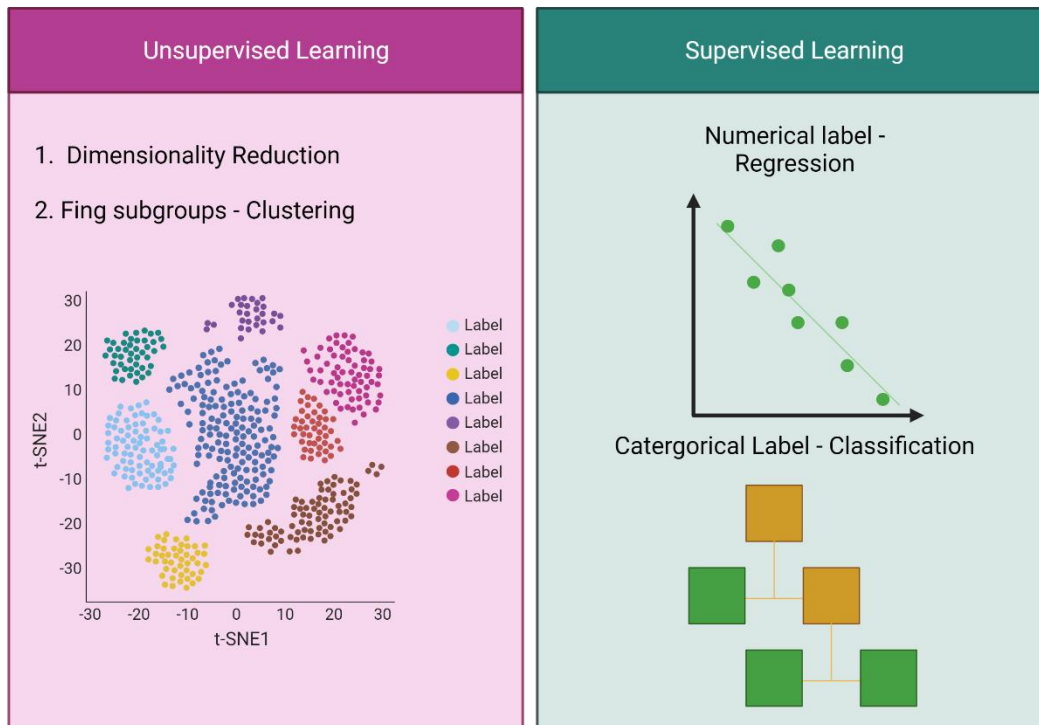


Figure 1. Machine learning methods can be classified into two general types: unsupervised and supervised. Unsupervised machine learning methods first undergo dimensionality reduction and then clustering in order to group like objects together, with the goal of finding general structure within data. Supervised machine learning methods can either be used to predict numerical labels or categorical labels; using labeled data supervised methods create a statistical relationship between samples and outputs in order to then predict one new data.

Unsupervised exploration of amphiphilic values

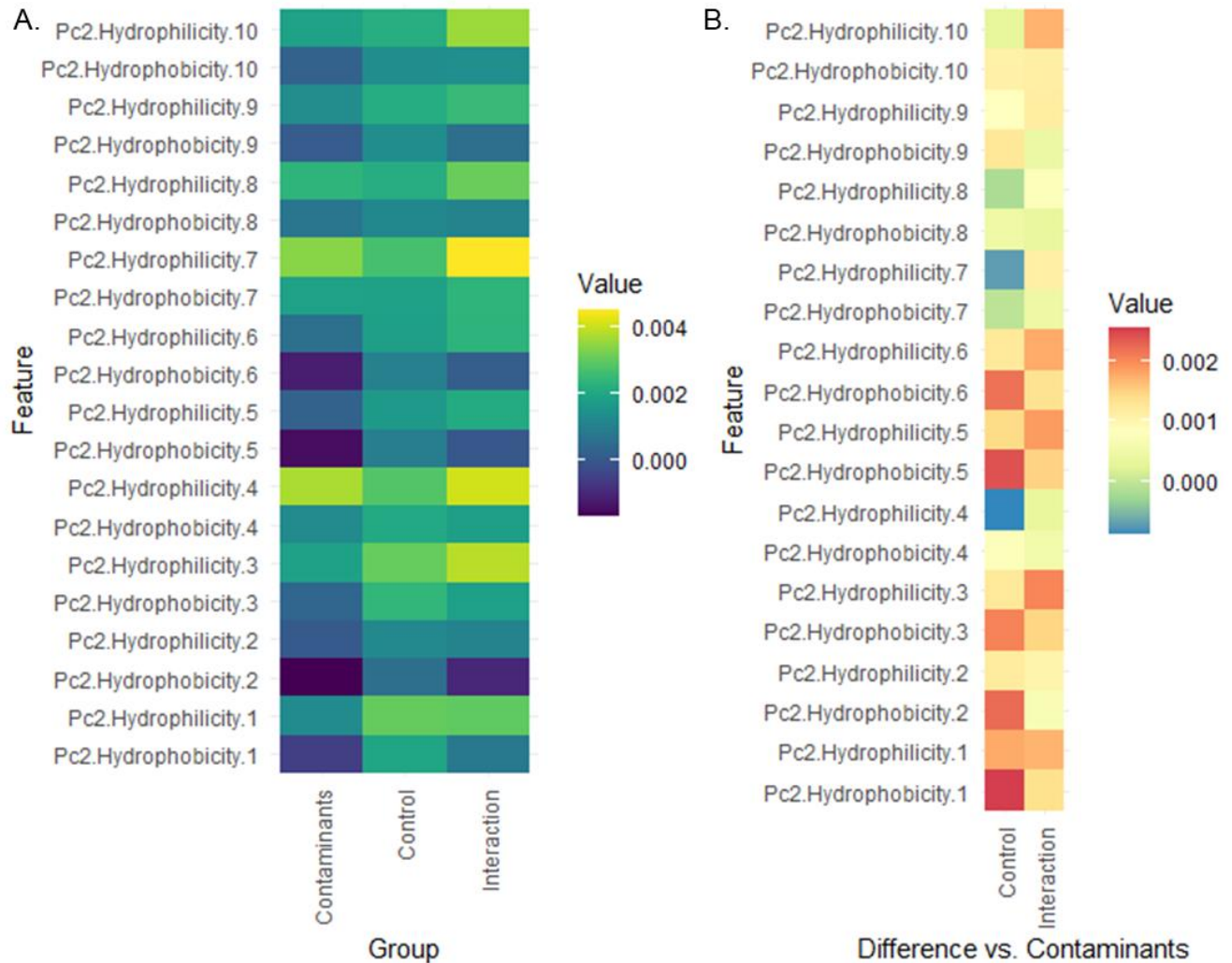


Figure 2. Average amphiphilicity differs across contaminants, control and interaction proteins. A. Amphiphilicity values were calculated across amino acids and averaged for all proteins found in BioID CRAPome contaminants purification, 1,000 random proteins found in the UNIPROT database, and BUB1 IntAct interaction proteins. B. Amphiphilic values for random control and BUB1 interaction proteins were compared to BioID contaminant proteins.

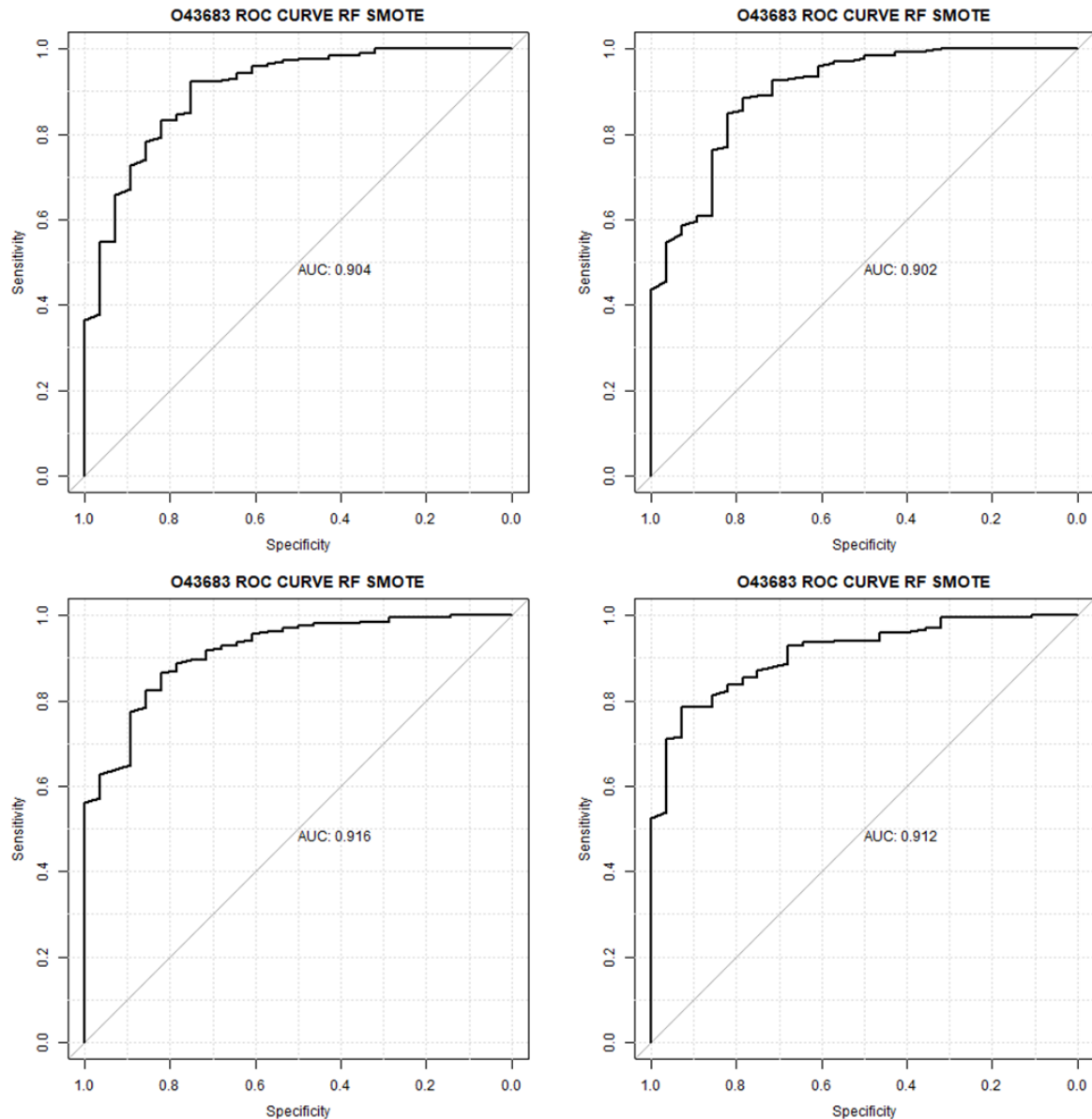


Figure 3. Supervised machine learning predicts contaminants in BioID experiments of spindle assembly checkpoint proteins. Amphiphilic features were extracted from CRAPome BioID contaminants, 1,000 random proteins found in the Uniprot database, and SAC protein interactions according to the IntAct database. Database was separated into two classes; BioID CRAPome contaminants and probable interactors which consisted of IntAct interactions and random Uniprot proteins. The datasets were then divided into a training (70%) and test set (30%) for each SAC component. Random forest modeling was used to create a model with the training set and accuracy was measured with test set. Accuracy was consistent across different SAC components.

REFERENCES

1. Apweiler, R., Bairoch, A., Wu, C.H., Barker, W.C., Boeckmann, B., Ferro, S., Gasteiger, E., Huang, H., Lopez, R., Magrane, M., Martin, M.J., Natale, D.A., O'Donovan, C., Redaschi, N., and Yeh, L.-S.L. (2004). UniProt: the Universal Protein knowledgebase. *Nucleic acids research* 32.
2. Arnold, R.J., Jayasankar, N., Aggarwal, D., Tang, H., and Radivojac, P. (2006). A machine learning approach to predicting peptide fragmentation spectra. *Pacific Symposium on Biocomputing*. Pacific Symposium on Biocomputing.
3. Badillo, S., Banfai, B., Birzele, F., Davydov, I.I., Hutchinson, L., Kam-Thong, T., Siebourg-Polster, J., Steiert, B., and Zhang, J.D. (2020). An Introduction to Machine Learning. *Clinical pharmacology and therapeutics* 107.
4. Blagus, R., and Lusa, L. (2013). SMOTE for high-dimensional class-imbalanced data. *BMC bioinformatics* 14.
5. Camacho, D.M., Collins, K.M., Powers, R.K., Costello, J.C., and Collins, J.J. (2018). Next-Generation Machine Learning for Biological Networks. *Cell* 173.
6. Chou, K.-C. (2005). Using amphiphilic pseudo amino acid composition to predict enzyme subfamily classes. *Bioinformatics (Oxford, England)* 21.
7. Chou, K.C. (2001). Prediction of protein cellular attributes using pseudo-amino acid composition. *Proteins* 43.
8. Demichev, V., Messner, C.B., Vernardis, S.I., Lilley, K.S., and Ralser, M. (2020). DIA-NN: neural networks and interference correction enable deep proteome coverage in high throughput. *Nature methods* 17.

9. Fuente, A.d.I. (2010). From 'differential expression' to 'differential networking' - identification of dysfunctional regulatory networks in diseases. *Trends in genetics* : TIG 26.
10. Henderson, A.R. (2005). The bootstrap: a technique for data-driven statistics. Using computer-intensive analyses to explore experimental data. *Clinica chimica acta; international journal of clinical chemistry* 359.
11. Hermjakob, H., Montecchi-Palazzi, L., Lewington, C., Mudali, S., Kerrien, S., Orchard, S., Vingron, M., Roechert, B., Roepstorff, P., Valencia, A., Margalit, H., Armstrong, J., Bairoch, A., Cesareni, G., Sherman, D., and Apweiler, R. (2004). IntAct: an open source molecular interaction database. *Nucleic acids research* 32.
12. Jin, L., Bi, Y., Hu, C., Qu, J., Shen, S., Wang, X., and Tian, Y. (2021). A comparative study of evaluating missing value imputation methods in label-free proteomics. *Scientific reports* 11.
13. Ma, C., Zhu, Z., Ye, J., Yang, J., Pei, J., Xu, S., Zhou, R., Yu, C., Mo, F., Wen, B., and Liu, S. (2017). DeepRT: deep learning for peptide retention time prediction in proteomics.
14. Mall, R., Cerulo, L., Bensmail, H., Iavarone, A., and Ceccarelli, M. (2017). Detection of statistically significant network changes in complex biological networks. *BMC systems biology* 11.
15. Mellacheruvu, D., Wright, Z., Couzens, A.L., Lambert, J.-P., St-Denis, N.A., Li, T., Miteva, Y.V., Hauri, S., Sardi, M.E., Low, T.Y., Halim, V.A., Bagshaw, R.D., Hubner, N.C., Al-Hakim, A., Bouchard, A., Faubert, D., Fermin, D., Dunham,

- W.H., Goudreault, M., Lin, Z.-Y., Badillo, B.G., Pawson, T., Durocher, D., Coulombe, B., Aebersold, R., Superti-Furga, G., Colinge, J., Heck, A.J.R., Choi, H., Gstaiger, M., Mohammed, S., Cristea, I.M., Bennett, K.L., Washburn, M.P., Raught, B., Ewing, R.M., Gingras, A.-C., and Nesvizhskii, A.I. (2013). The CRAPome: a contaminant repository for affinity purification-mass spectrometry data. *Nature methods* 10.
16. Meyer, J.G. (2021). Deep learning neural network tools for proteomics. *Cell Reports Methods* 1, 100003.
17. Oughtred, R., Stark, C., Breitkreutz, B.-J., Rust, J., Boucher, L., Chang, C., Kolas, N., O'Donnell, L., Leung, G., McAdam, R., Zhang, F., Dolma, S., Willems, A., Coulombe-Huntington, J., Chatr-Aryamontri, A., Dolinsk, K., and Tyers, M. (2019). The BioGRID interaction database: 2019 update. *Nucleic acids research* 47.
18. Petritis, K., Kangas, L.J., Ferguson, P.L., Anderson, G.A., Pasa-Tolić, L., Lipton, M.S., Auberry, K.J., Strittmatter, E.F., Shen, Y., Zhao, R., and Smith, R.D. (2003). Use of artificial neural networks for the accurate prediction of peptide liquid chromatography elution times in proteome analyses. *Analytical chemistry* 75.
19. Rajkomar, A., Dean, J., and Kohane, I. (2019). Machine Learning in Medicine. *The New England journal of medicine* 380.
20. Sagawa, C.H.D., Zaini, P.A., Assis, R.d.A.B., Saxe, H., Salemi, M., Jacobson, A., Wilmarth, P.A., Phinney, B.S., and Dandekar, A.M. (2020). Deep Learning Neural

- Network Prediction Method Improves Proteome Profiling of Vascular Sap of Grapevines during Pierce's Disease Development. *Biology* 9.
21. Shen, D., Wu, G., and Suk, H.-I. (2017). Deep Learning in Medical Image Analysis. *Annual review of biomedical engineering* 19.
22. Spivak, M., Weston, J., Tomazela, D., MacCoss, M.J., and Noble, W.S. (2012). Direct maximization of protein identifications from tandem mass spectra. *Molecular & cellular proteomics : MCP* 11.
23. Tiwary, S., Levy, R., Gutenbrunner, P., Soto, F.S., Palaniappan, K.K., Deming, L., Berndl, M., Brant, A., Cimermancic, P., and Cox, J. (2019). High-quality MS/MS spectrum prediction for data-dependent and data-independent acquisition data analysis. *Nature methods* 16.
24. Wang, C., Wang, W., Lu, K., Zhang, J., Chen, P., and Wang, B. (2020). Predicting Drug-Target Interactions with Electrotopological State Fingerprints and Amphiphilic Pseudo Amino Acid Composition. *International journal of molecular sciences* 21.
25. Xiao, N., Xu, Q.-S., and Cao, D.-S. (2021). CRAN - Package protr.
26. Zhang, G.-Y., and Fang, B.-S. (2008). Predicting the cofactors of oxidoreductases based on amino acid composition distribution and Chou's amphiphilic pseudo-amino acid composition. *Journal of theoretical biology* 253.
27. Zhou, X.-X., Zeng, W.-F., Chi, H., Luo, C., Liu, C., Zhan, J., He, S.-M., and Zhang, Z. (2017). pDeep: Predicting MS/MS Spectra of Peptides with Deep Learning. *Analytical chemistry* 89.

CANVS Instructions

Getting started with CANVS

The goal of this document is to get you started with CANVS as quickly as possible.

CANVS is an open-access easy-to-use pipeline for studying protein-protein interaction/association data. CANVS was created so that scientists with no computational background can quickly analyze mass spectrometry data from affinity-based and proximity-based protein purifications, with an emphasis on identifying biologically interesting protein interactions/associations that can be further validated and explored. The pipeline can be divided into the following steps:

1. Uploading data
2. Cleaning data
3. Analysis with system biology databases/Network visuals

The tabs in the application reflect these steps, where the network visuals steps can be done with or without filtering the results with GO terms. Each step is described in more detail below.

0. Download and Setup

CANVS is packaged as an R Shiny dashboard app but since it includes data from large databases, the application is deployed locally. Currently the local version of the app can only be used in a windows systems. To install CANVS, download the CANVS.zip file from (<https://sourceforge.net/projects/canvs/files/>). Unzip the file in a location where you want to save the application. Open the folder and right click on the CANVS.bat file, then select create a shortcut. Name the shortcut CANVS and then drag the shortcut to the desktop. Then double click on the new shortcut and CANVS will open in a web browser window. Please make sure that there are no antivirus software blocking application connections to ports in the computer and that firewall does not block use of ports.

1. Uploading Data

CANVS accepts multiple csv or tab delimited files as long as each data table has the following information, along with these exact column names;

- *Accession* - UNIPROT accession ID.
- *Description* - Description of the protein identified.
- *File* - Unique filename for the purification.
- *Bait* - The bait or protein that was affinity tagged. This is usually the protein of interest that is being studied.

If the user chooses to carry out a semi-quantitative qualitative analysis the user should also include a score column that represents a quantitative value at the protein level.

- *Score* - a quantitative score at the protein level.

Data is uploaded in the “upload” tab, however it’s important to remember that the pipeline will only run if the above columns are included in the data.

Instructions of acceptable data structures are summarized in the first box “1. Instructions”. Data is uploaded in the second box “2. Upload”. Multiple files can be uploaded in each of the control and experimental sections. Once the *Upload Data* button is pressed, the results will be rendered as data tables in the final two boxes, “3. Control” and “4. Experimental”. Each box can be expanded in order to view the data tables with the uploaded control and experimental data. Additionally, the user has the option of uploading a single file for either the control or experimental as long as it is formatted with the columns above and replicates are indicated with unique file names in File column.

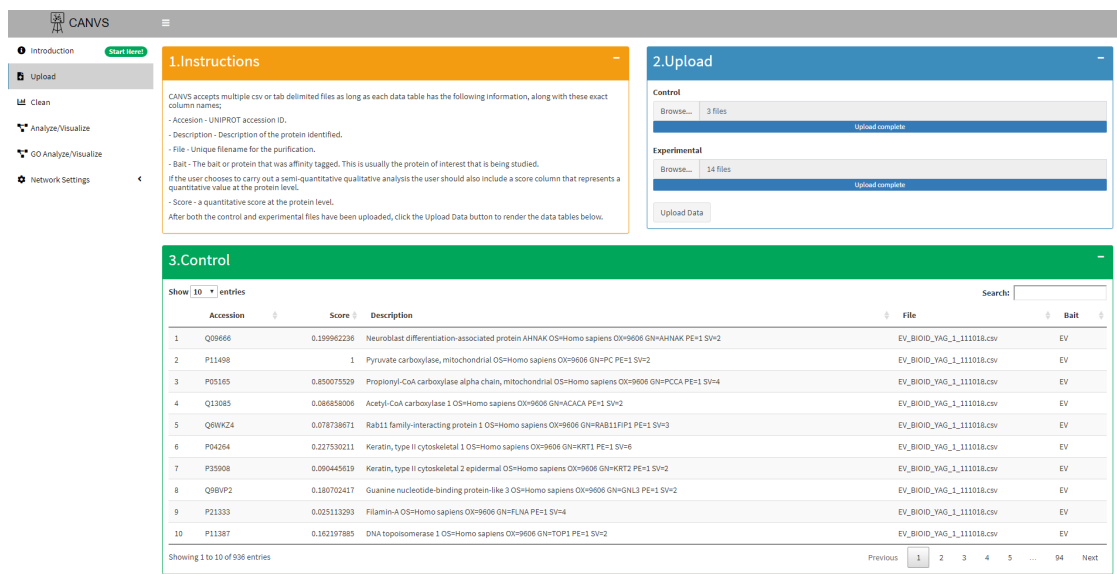


Figure S1: Upload Tab

2. Cleaning Data

Once the data has been uploaded, the user has to set up the analysis method for the pipeline to run. There are three main settings that the user has to consider:

- Perform Analysis - whether to perform analysis or not.
- Protein Replicates - Filter proteins to consider only proteins that appear in n replicates.
- Normalization - Normalize score values by the median across purifications.

CANVS can perform log2fold change and a two-tailed t-test. Coupled together it will render a volcano plot, however the user also has the option of just using log2fold change as a metric to filter data. Once the analysis is complete the user can change settings like the log fold value to filter by, the P value to consider or whether to look at UP or DOWN regulated proteins. The user also has the option of not performing any analysis or using already processed data as input

Each item is described in more detail below.

Statistical Analysis

Once the data is filtered by replicates and normalized, the user has the option of applying statistical tests to proteins shared in the control and experimental proteins.

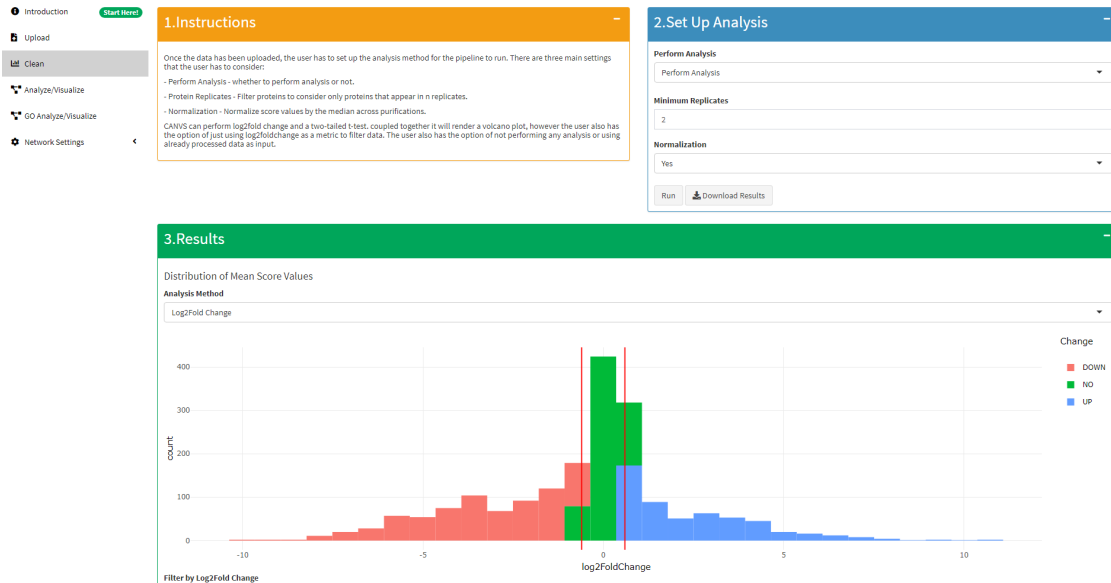


Figure S2: Clean Tab

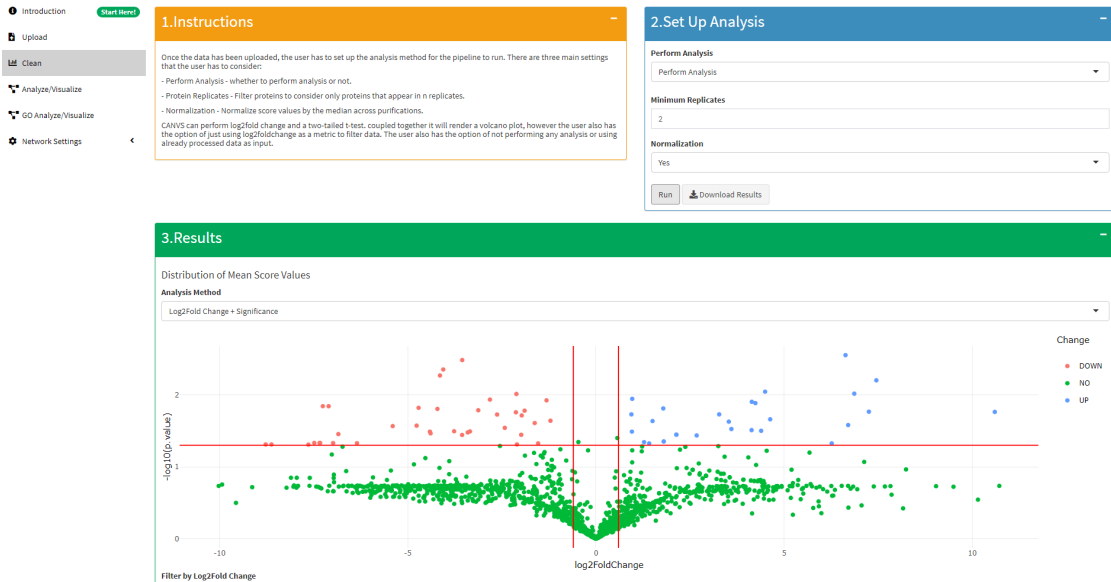


Figure S3: Clean Tab

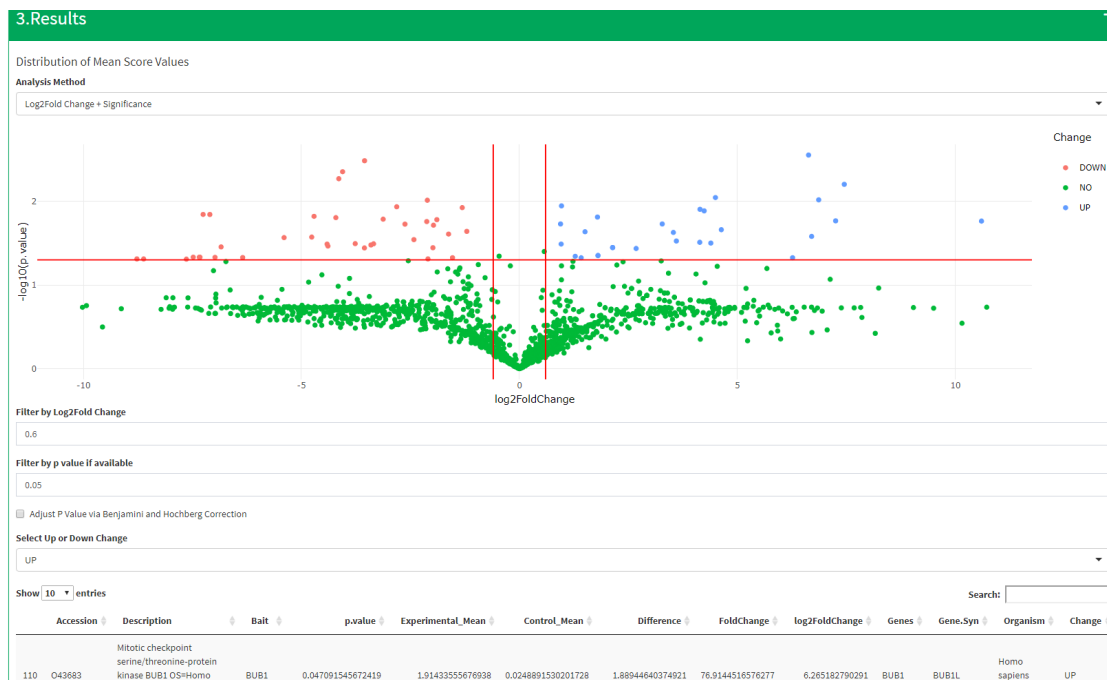


Figure S4: Clean Tab

A. Perform Analysis - CANVS first calculates the difference between the logarithmized mean protein intensities between experimental and control purifications. The log base-2 of the fold change is then calculated since it is beneficial to represent the distribution around zero, a value that indicates no change for a protein between the experimental and control. If a large number of replicates (we recommend three biological replicates and two technical replicates) are used in the analysis, the user has the option of also calculating the significance in the log2fold change using a two-tailed t-test. Comparing the negative log of the P value to the log2fold change creates a volcano plot where background proteins cluster at zero.

B. No Analysis - All proteins in the experimental are considered hits.

For the first option experimental and control data sets must be supplied, however if the user chooses not to carry out a statistical analysis only experimental data is needed. No analysis might be useful if the user has done a quantitative analysis somewhere else and wants to use CANVS for analysis/visual purposes.

Filtering proteins by replicates

The user has the option to filter proteins by how many times they were identified in multiple replicates. This is particularly helpful to experimental set ups that included multiple replicates and want to increase confidence by only including proteins found in n replicates in the results. Default setting is set to 2 but can be changed using the drop down menu.

Normalizing

Currently, CANVS supports normalization by the median. Briefly, CANVS scales samples so that each purification has the same median value.

Results

Results can then be viewed in the box below and the user can download them in a csv format by clicking the Download Results button. The result table will also be used in the rest of the pipeline to create network visuals. The table and analysis are reactive to user inputs and so if the user chooses to analyze data in a different method, the user would first update results in the analysis tab and then run the rest of the pipeline. The filtering tools allow the user to change the conditions by which to consider hits. The user has the option of filtering either by log2fold change or log2fold change in combination with a given P value. Additionally the user can correct P values by using the Benjamini and Hochberg correction. Briefly, adjusting the P value controls the false discovery rate and therefore corrects for the expected amount of false positives among all positives which rejected the null hypothesis.

3. Analysis/Visualization

The application features 2 final tabs that provide the user with visual network representations of the results. The first tab creates networks for all identified proteins while the second tab features a Gene Ontology based filtering tool that allows the user to create visual representations of proteins with certain GO terms. Both tabs feature annotated data in the form of protein-protein and protein-complex networks. To incorporate protein-protein information we integrated the Biological General Repository for Interaction Datasets (BioGRID v. 3.5). To incorporate protein-complex information, we integrated the Comprehensive Resource of Mammalian Protein Complexes (CORUM v. 3.0).

Protein Networks Considering All Hits

The first visualization tab features 4 different ways to analyze results:

- Bait-Bait Interaction/Association Networks
- Protein-Protein Interaction/Association Networks
- CORUM Complex Protein-Protein Interaction/Association Networks
- BioGRID Protein-Protein Interaction/Association Networks

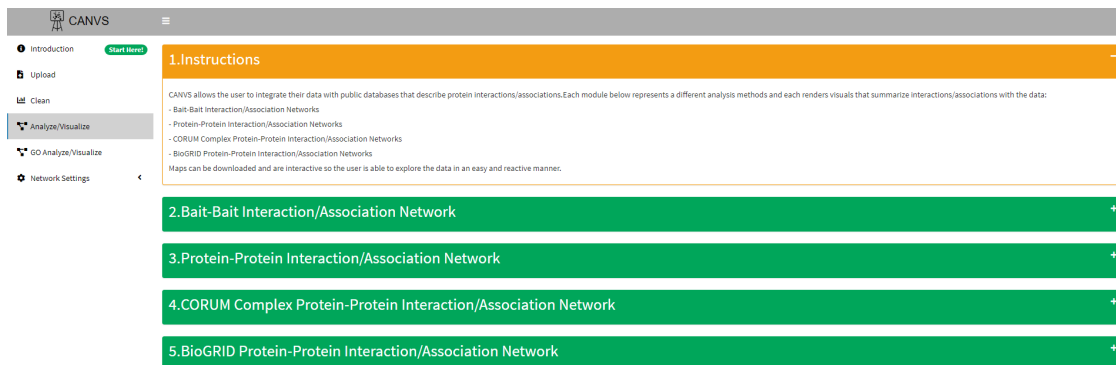


Figure S5: Main Analyze/Visualize Tab

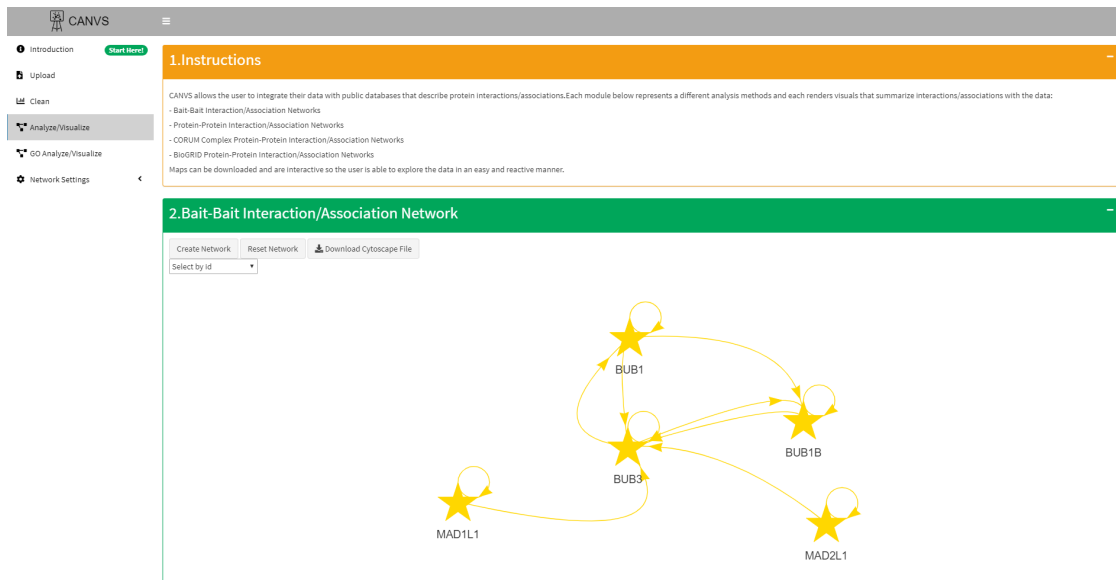


Figure S6: Main Analyze/Visualize Expanded Tab

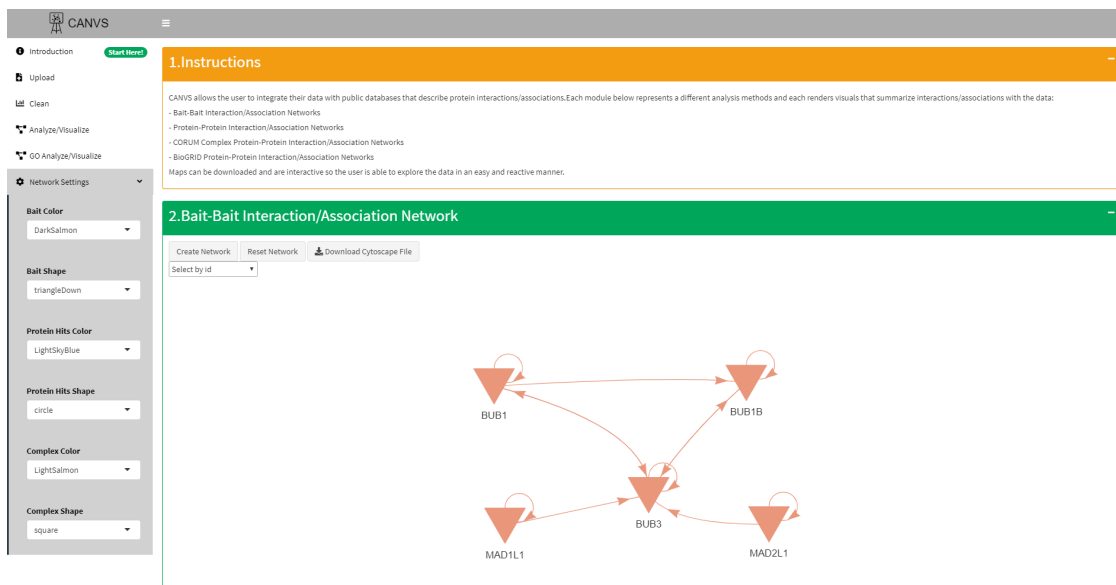


Figure S7: Change Color and Shape in Networks

Users can create networks corresponding to the network type of that box by pressing the + button and expanding each box. For example, a bait-bait network diagram is shown below after expanding the first box.

All networks are customizable (in terms of color and shape of nodes) by clicking the Network Settings link in the sidebar. For example, the color and shape of the nodes in the previous network are changed from gold to salmon and from star to down triangle by changing the color and shape of baits in the sidebar, resetting the network, and then creating the network again. All networks carry the same logic and settings in terms of color and shape to make the analysis consistent throughout tabs.

To reset a network after applying any change: 1. Click Reset Network 2. Click Create Network

Furthremore, the user has the option of downloading Cytoscape network files that can be used to create network visuals in Cytoscape. To do so press the Download Cytoscape File button in each network and then upload the file to Cytoscape as a network.

Filter by GO Terms

CANVS also allows the user to filter protein results by GO terms associated with the proteins in the results. To do so, first search GO Terms by a keyword which will be used to search for any GO terms associated with the results that have that keyword. The terms that have the keyword and the subterms, or child terms in Gene Ontology, will be selected. These will be summarized in a table and the association of how the terms might be related to each other can also be analyzed in a network. The network will color code GO terms based on whether they are biological processes, molecular functions, or cellular components (the three categories Gene Ontology uses to describe GO terms).

1. GO Based Analysis

CANVS also allows the user to filter protein results by GO terms associated with the proteins in the results. To do so, first search GO Terms by a keyword which will be used to search for any GO terms associated with the results that have that keyword. The terms that have the keyword and the subterms, or child terms in Gene Ontology, will be selected. These will be summarized in a table and the association of how the terms might be related to each other can also be analyzed in a network map. The network map will color code GO terms based on whether they are biological processes, molecular functions, or cellular components (the three categories Gene Ontology uses to describe GO terms). To carry out this analysis:

A. Search by keyword and select Geneontologies of interest in table in box 2. Terms of interest are selected by clicking on each term in the table. If the user wants to select all terms, then no terms need to be selected CANVS will automatically select all terms.

B. Create Networks. The rest of the networks in the tab are:

- Protein-Protein Interaction/Association Networks
- CORUM Complex Protein-Protein Interaction/Association Networks
- BioGRID Protein-Protein Interaction/Association Networks

CANVS will then only select the proteins associated with the GO terms that the user picked (or all the GO terms in the search if the user did not select any GO terms). The environment is reactive and so it's possible to add or take away terms and then recreate the networks.

2. Select GO Terms

Search by Keyword

mitotic

Key:

- Biological Process
- Molecular Function
- Cellular Component

Select by id

Search Create Network Reset Network Download Cytoscape File

attachment of mitotic spindle microtubules to kinetochores

mitotic metaphase plate congression

mitotic chromosome movement towards spindle pole

mitotic sister chromatid segregation

mitotic sister chromatid separation

Figure S8: GO Analyze/Visualize Tab

Terms of interest are selected by clicking on each term in the table. If the user wants to select all terms, then no terms need to be selected CANVS will automatically select all terms. The rest of the networks in the tab are:

- Protein-Protein Interaction/Association Networks
- CORUM Complex Protein-Protein Interaction/Association Networks
- BioGRID Protein-Protein Interaction/Association Networks

CANVS will then only select the proteins associated with the GO terms that the user picked (or all the GO terms in the search if the user did not select any GO terms). The environment is reactive and so it's possible to add or take away terms and then recreate the networks.

2. Select GO Terms

Search by Keyword

● Biological Process
● Molecular Function
● Cellular Component

Select by id ▼

Show 10 entries

	GOID	Term	Ontology
4	GO:0090307	mitotic spindle assembly	BP
8	GO:0072686	mitotic spindle	CC
954	GO:1900087	positive regulation of G1/S transition of mitotic cell cycle	BP
1223	GO:0000086	G2/M transition of mitotic cell cycle	BP
1228	GO:0010389	regulation of G2/M transition of mitotic cell cycle	BP
1854	GO:0097431	mitotic spindle pole	CC
1917	GO:1901990	regulation of mitotic cell cycle phase transition	BP
2669	GO:0045930	negative regulation of mitotic cell cycle	BP
2674	GO:0007094	mitotic spindle assembly checkpoint	BP
2675	GO:0007093	mitotic cell cycle checkpoint	BP

Showing 1 to 10 of 32 entries

Previous
1
2
3
4
Next

Figure S9: Select GO Terms

CORUM Networks

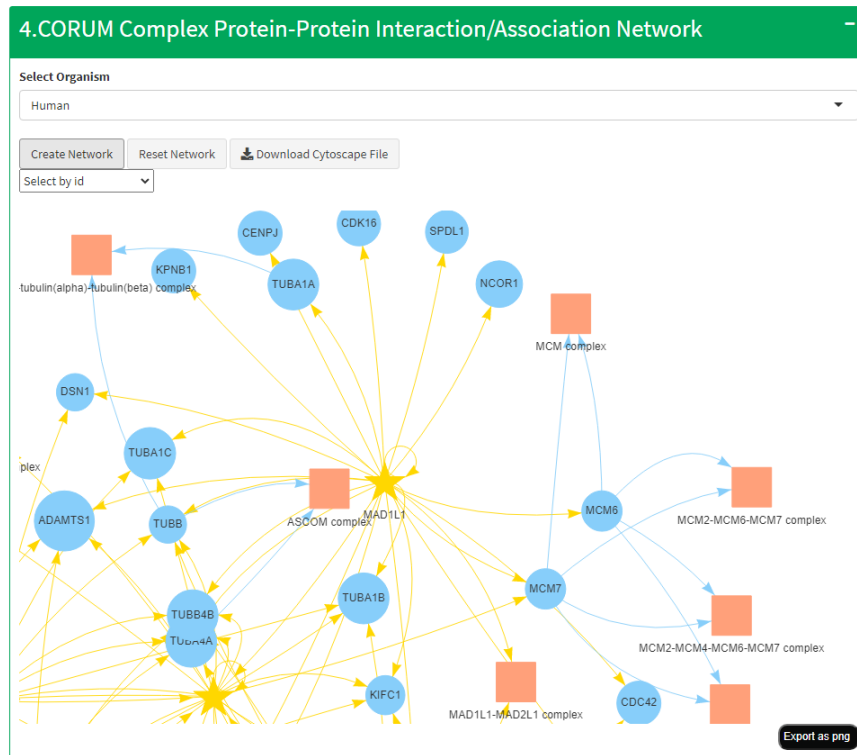


Figure S10: CORUM Example

The user can create CORUM (CORUM v. 3.0) based networks with proteins from the general results or proteins pertaining to particular GO Terms. To expand the CORUM Network box, select an organism of interest and click create network. This will render a network with PPI/As and their respective complexes under the organism chosen. To create an informative network, only complexes that are shared by two or more proteins are considered when making the network. Currently CORUM supports the following organisms:

- Human
- Mouse
- Pig
- Bovine
- Rat
- Mammalia
- Rabbit
- Dog
- Hamster
- Mink

The user has the option of changing the original search terms or updating the results, which results in a different network that the user can render by first clicking reset network and then creating the network again. Below is an example:

BioGRID Networks

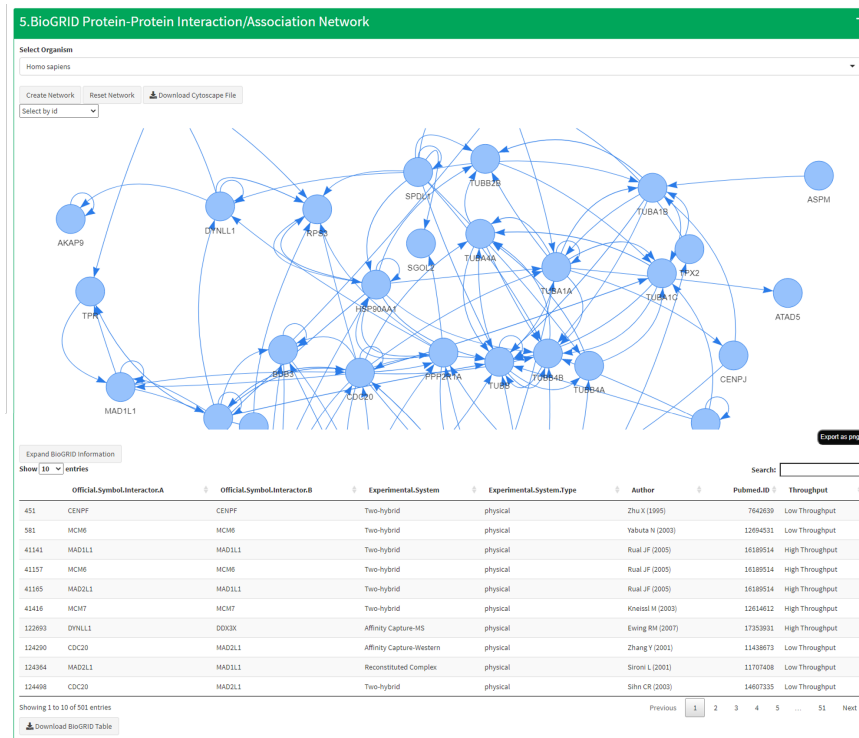


Figure S11: BioGRID Example

To contextualize the results of PPI/A data with previously known interaction data, CANVS integrates BioGRID (BioGRID V. 3.5). To create these networks, click on the BioGRID network box and search for an organism of interest. Click on create network which will render a network showing how the results are related in the context of BioGRID interactions. The user also has the option to expand BioGRID information and create a datatable that summarizes each interaction and offers the user information about its nature and where it was initially reported. This datatable can also be downloaded. BioGRID currently supports the following organisms:

- Homo sapiens
- Drosophila melanogaster
- Caenorhabditis elegans
- Saccharomyces cerevisiae S288C
- Schizosaccharomyces pombe 972h-
- Mus musculus
- Rattus norvegicus
- Canis lupus familiaris
- Arabidopsis thaliana
- Bos taurus
- Gallus gallus
- Escherichia coli BW2952
- Escherichia coli str. K-12 substr. MG1655
- Neurospora crassa OR74A
- Dictyostelium discoideum AX4
- Human papillomavirus type 16
- Escherichia coli str. K-12 substr. W3110

- *Ustilago maydis* 521
- *Candida albicans* SC5314
- *Aspergillus nidulans* FGSC A4
- *Bacillus subtilis* subsp. *subtilis* str. 168
- *Anopheles gambiae* str. PEST
- *Pediculus humanus corporis*
- *Selaginella moellendorffii*
- *Escherichia coli* K-12
- *Mycobacterium tuberculosis* H37Rv
- *Chlorocebus sabaues*
- *Oryza sativa* Japonica Group
- Human gammaherpesvirus 8
- *Plasmodium falciparum* 3D7
- Human betaherpesvirus 6B
- Human betaherpesvirus 6A
- *Vitis vinifera*
- Tobacco mosaic virus
- Simian immunodeficiency virus
- Human immunodeficiency virus 2
- Human immunodeficiency virus 1
- Hepacivirus C
- Simian virus 40
- Human gammaherpesvirus 4
- Human betaherpesvirus 5
- Human alphaherpesvirus 3
- Human alphaherpesvirus 2
- Human alphaherpesvirus 1
- *Vaccinia virus*
- *Cavia porcellus*
- *Cricetulus griseus*
- *Oryctolagus cuniculus*
- *Sus scrofa*
- *Equus caballus*
- *Pan troglodytes*
- *Macaca mulatta*
- *Meleagris gallopavo*
- *Xenopus laevis*
- *Danio rerio*
- *Strongylocentrotus purpuratus*
- *Apis mellifera*
- *Zea mays*
- *Solanum tuberosum*
- *Nicotiana tomentosiformis*
- *Solanum lycopersicum*
- *Ricinus communis*
- *Glycine max*
- *Chlamydomonas reinhardtii*

The user has the option of changing the original search terms or updating the results which would result in a different network which the user can render by first clicking reset network and then creating the network again. Below is an example of the box:

4. Conclusion

CANVS is meant to be used as an interactive platform to explore how proteins from PPI/A experiments might be associated with each other. Proper experimental validation of PPI/As is then required. Furthermore, as proteinformatic databases expand to include data from more organisms, CANVS can be extended to include those organisms. Nonetheless, CANVS empowers any scientist to carry out a simple straight-forward bioinformatic analysis of PPI/A data without having the technical/coding skills to do so.

SUPPLEMENTAL MATERIAL

Mapping Proximity Associations of Core Spindle Assembly Checkpoint Proteins

Yenni A. Garcia¹, Erick F. Velasquez¹, Lucy W. Gao², Ankur A. Gholkar¹, Kevin M. Clutario¹, Keith Cheung¹, Taylor Williams-Hamilton¹, Julian P. Whitelegge^{2,3,4} and Jorge Z. Torres^{1,3,4*}

¹Department of Chemistry and Biochemistry, University of California, Los Angeles, CA 90095, USA

²Pasarow Mass Spectrometry Laboratory, The Jane and Terry Semel Institute for Neuroscience and Human Behavior, David Geffen School of Medicine, University of California, Los Angeles, CA 90095, USA

³Molecular Biology Institute, University of California, Los Angeles, CA 90095, USA

⁴Jonsson Comprehensive Cancer Center, University of California, Los Angeles, CA 90095, USA

*Corresponding author:

Jorge Z. Torres
607 Charles E. Young Drive East
Los Angeles, CA 90095
Phone: 310-206-2092
Fax: 310-206-5213
torres@chem.ucla.edu

Table of Contents

- Figure S1.** Generation of BioID2-tagged SAC protein inducible stable cell lines.
- Figure S2.** Workflow of mass spectrometry data acquisition and analysis.
- Figure S3.** Establishment of inducible BioID2-tagged SAC protein (BUB1, BUB3, BUBR1, MAD1L1 and MAD2L1) stable cell lines.
- Figure S4.** BioID2 biochemical purifications.
- Figure S5.** Core SAC protein proximity association maps.
- Figure S6.** Core SAC protein proximity network.
- Figure S7.** The BioID2-BUB1 proximity protein association map.
- Figure S8.** The BioID2-BUB3 proximity protein association map.
- Figure S9.** The BioID2-BUBR1/BUB1B proximity protein association map.
- Figure S10.** The BioID2-MAD1L1 proximity protein association map.
- Figure S11.** The BioID2-MAD2L1 proximity protein association map.
- Figure S12.** The core SAC protein proximity association network using kinetochore Gene Ontology annotations.
- Figure S13.** The core SAC protein proximity association network using mitotic spindle related Gene Ontology annotations.
- Figure S14.** The core SAC protein proximity association network using centromere related Gene Ontology annotations.
- Figure S15.** Kinetochore protein enrichment analysis.
- Figure S16.** The core SAC protein proximity association network using nuclear pore related Gene Ontology annotations.
- Figure S17.** Uncropped immunoblots and autoradiography blots for all figures.
- Figure S18.** Uncropped immunoblots for all supplemental figures.
- Table S1.** List of reagents used. (.xlsx)
- Table S2.** List of primers used. (.xlsx)
- Table S3.** List of vectors generated. (.xlsx)
- Table S4.** Summary of all identified peptides from all BioID2 purifications. (.xlsx)
- Table S5.** Summary of all identified proteins from all BioID2 purifications. (.xlsx)
- Table S6.** Summary of peptides for all proteins that were identified with one peptide sequence. (.xlsx)
- Table S7.** Summary of significant SAC protein proximity associated proteins. (.xlsx)
- Table S8.** List of Gene Ontology (GO) annotations used in the core SAC protein proximity association network analyses. (.xlsx)

Supplemental Figures

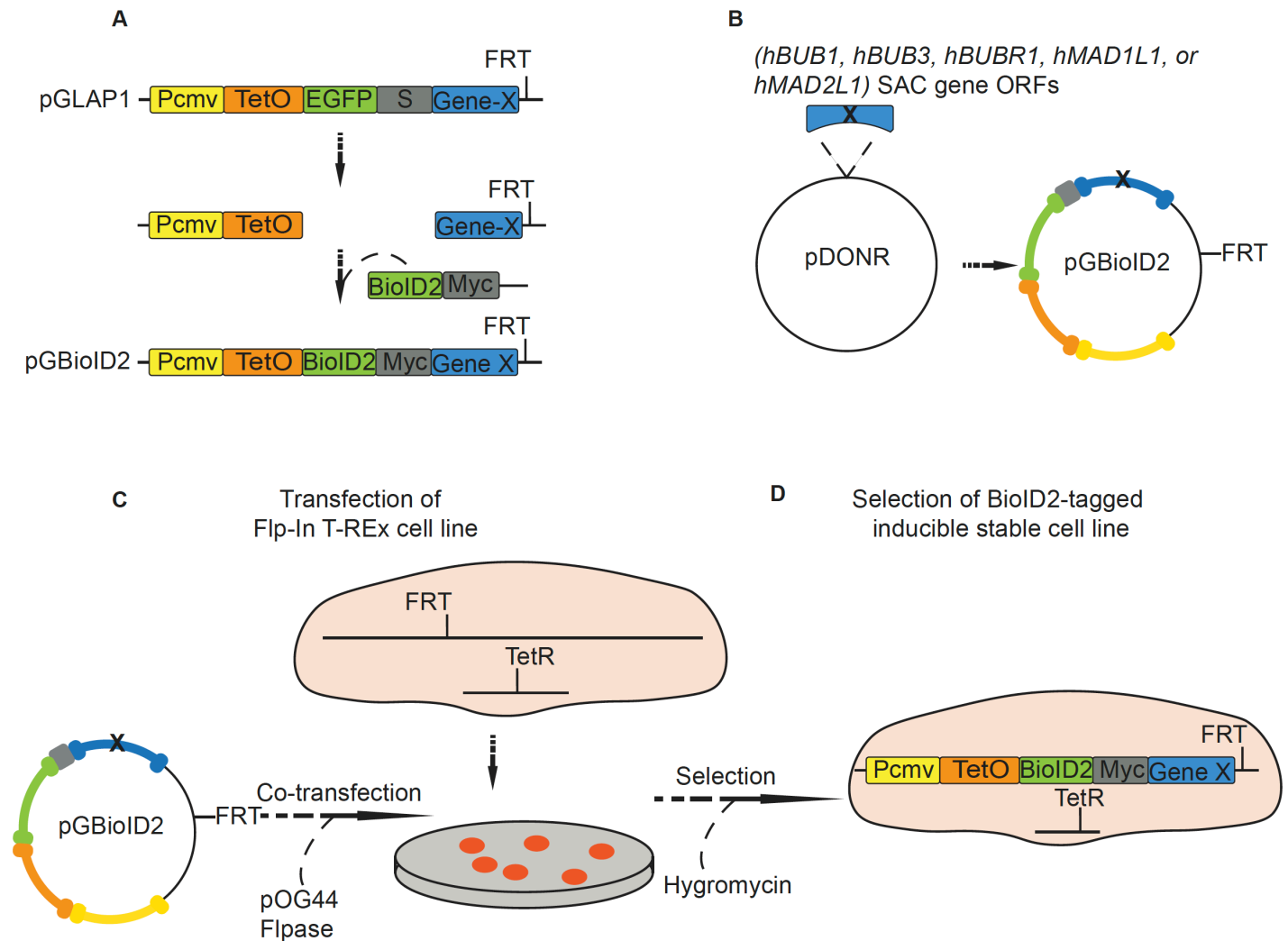


Figure S1. Generation of BioID2-tagged SAC protein inducible stable cell lines. (A) Generation of the Gateway compatible pGBioID2 vectors (pGBioID2-27 and pGBioID2-47) using the pGLAP1 vector backbone. S denotes S-tag and M denotes Myc-tag. (B) Transfer of core human SAC gene ORFs from the pDONR221 vector to the pGBioID2 vector. (C-D) pGBioID2-SAC gene vectors were co-transfected with the Flpase containing vector pOG44 into HeLa Flp-In-TREx cells and stable integrants were selected with Hygromycin. See Tables S2 and S3 for primer amplification sequences and information on the vectors generated.

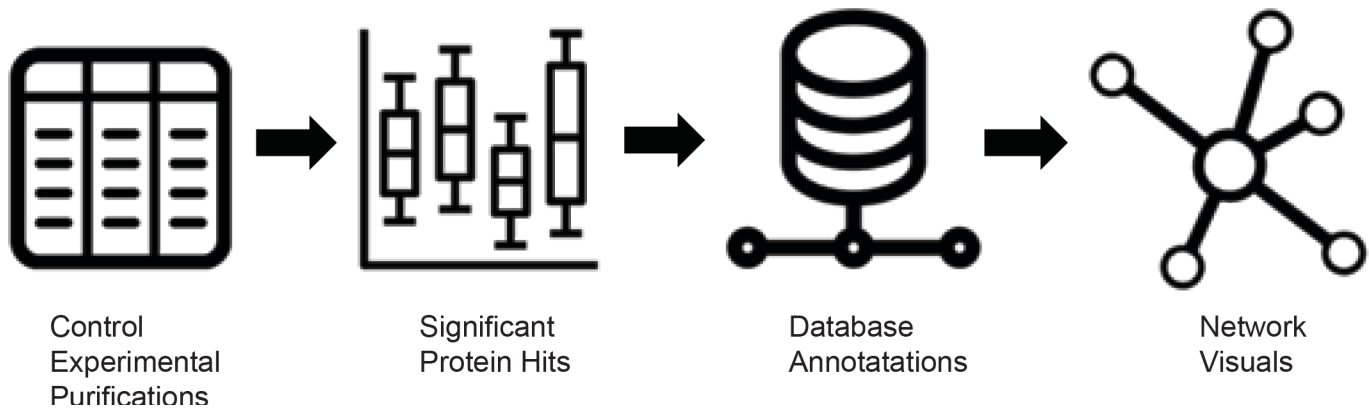


Figure S2. Workflow of mass spectrometry data acquisition and analysis. Control and experimental SAC core protein BioID2 purifications were analyzed by mass spectrometry and the mass spectrometry data was analyzed with a Mascot search that yielded 8,403 unique peptides, from which 903 unique proteins were identified. The experimental purifications, considering all 5 baits and replicates, identified 870 proteins while the control purifications identified 480 proteins. Due to the overlap between proteins identified in the control and experimental purifications, we determined hits by identifying proteins that showed a significant difference when compared to the control. Briefly, proteins shared between control and experimental purifications were compared by calculating the mean difference and then simulating the distribution of the mean difference (10,000 simulations). Once the distribution was simulated the 95% confidence interval (CI) were calculated and proteins which lied outside the 95% CI and also showed a higher spectral count in the experimental compared to the control were considered hits. A total of 378 unique proteins were considered hits across all experimental purifications. Protein hits were then queried in several databases (Geneontology; BioGRID; Reactome) and Cytoscape protein association networks were used to visualize the results.

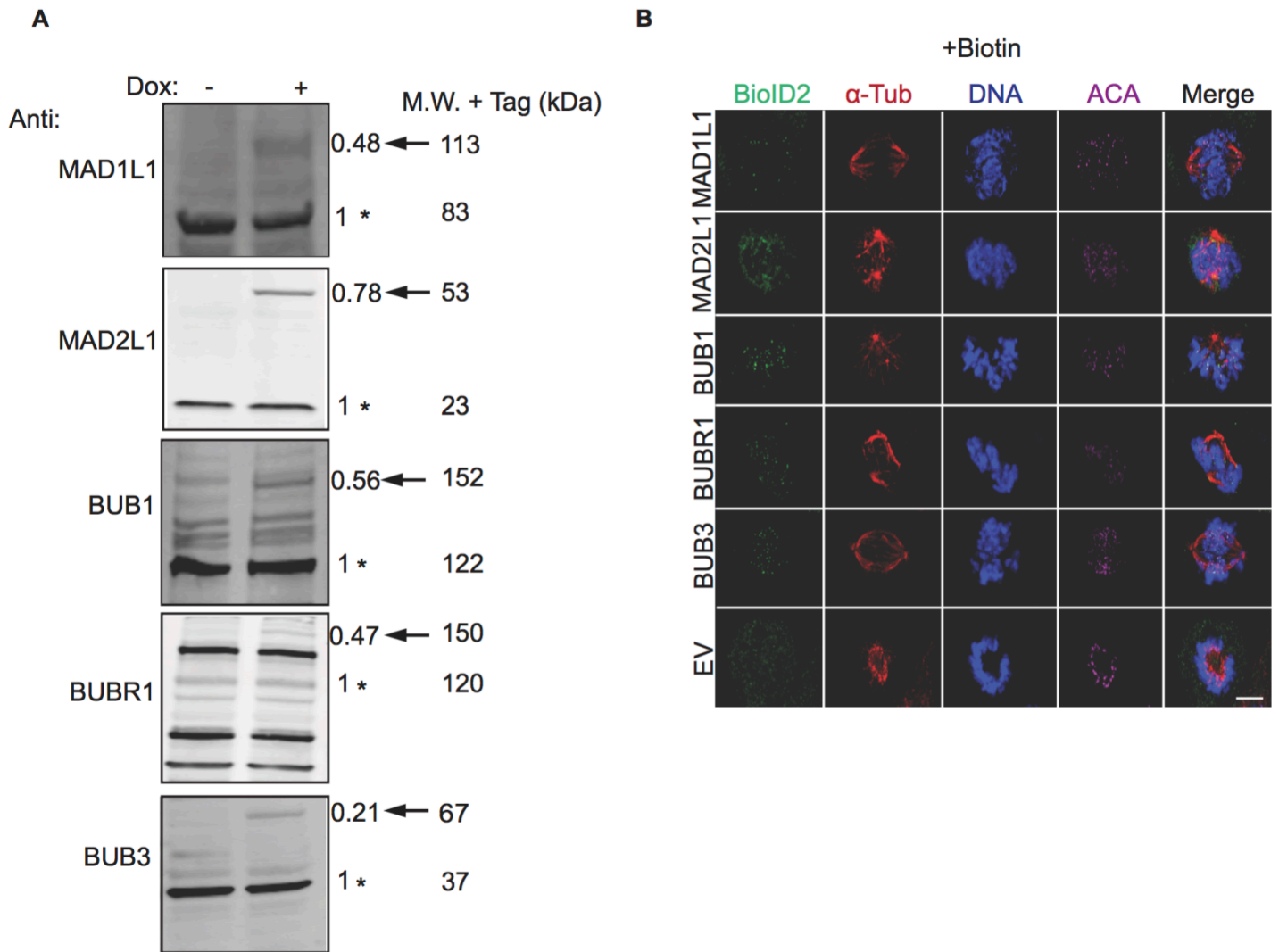


Figure S3. Establishment of inducible BioID2-tagged SAC protein (BUB1, BUB3, BUBR1, MAD1L1 and MAD2L1) stable cell lines. (A) Immunoblot analysis of extracts from doxycycline (Dox)-inducible BioID2-tag alone or BioID2-tagged SAC protein (BUB1; BUB3; BUBR1; MAD1L1; MAD2L1) expression cell lines in the absence (-) or presence (+) of Dox for 16 hours. For each cell line, blots were probed with the indicated antibodies that recognize the endogenous protein (labeled with a 1* to indicate 1x endogenous protein levels) and the corresponding overexpressed BioID2-tagged protein (labeled with .21-.78 and an arrow to indicate that their level of expression is below endogenous levels). M.W. indicates molecular weight. Note that BioID2-tagged SAC proteins are only expressed in the presence of Dox. (B) Fixed-cell immunofluorescence microscopy of the BioID2-tag alone (empty vector, EV) or the indicated BioID2-tagged SAC proteins during prometaphase, a time when the SAC is active. HeLa BioID2-tagged protein expression cell lines were induced with Dox and treated with 50 μ M Biotin for 16 hours, fixed and stained with Hoechst 33342 DNA dye and anti-BioID2, anti- α -Tubulin and anti-centromere antibodies (ACA). Bar indicates 5 μ m. Note that all BioID2-tagged SAC proteins localize to the kinetochore region, overlapping with the ACA signal, whereas the BioID2-tag alone (EV) was diffused throughout the cell.

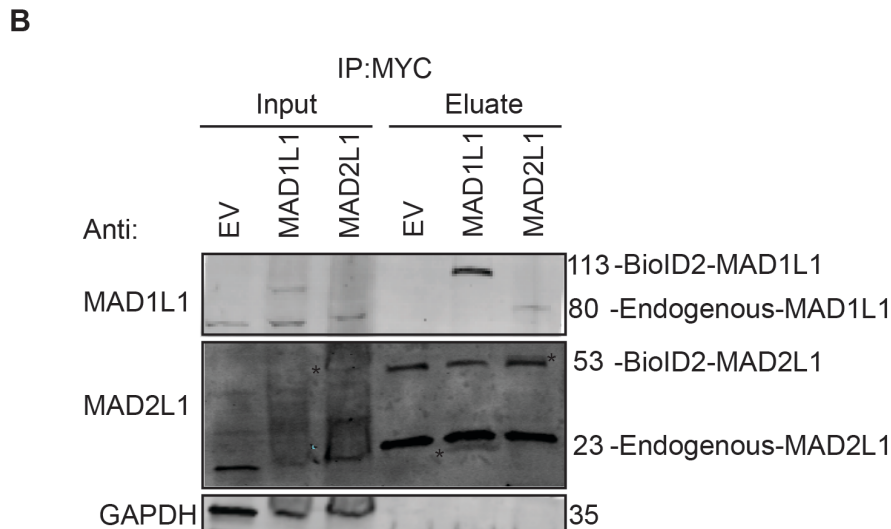
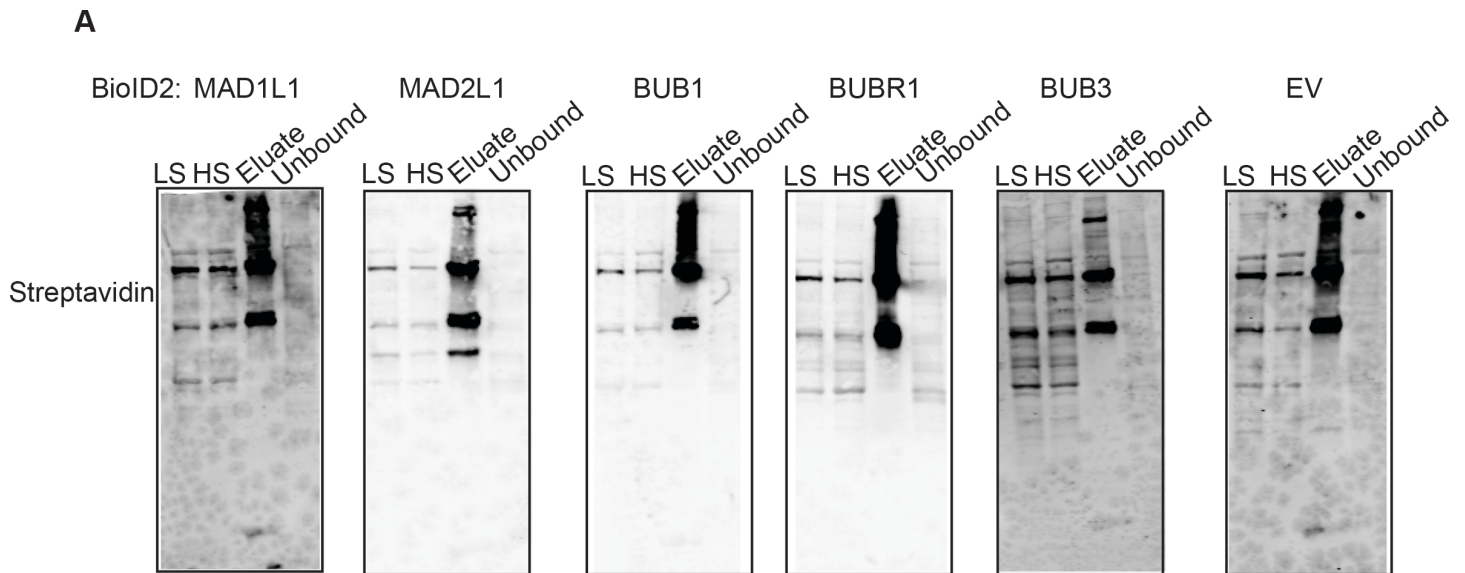


Figure S4. BioID2 biochemical purifications. (A) HeLa BioID2-tag alone (empty vector, EV) or BioID2-tagged SAC protein (BUB1; BUB3; BUBR1; MAD1L1; MAD2L1) stable cell lines were induced to express the BioID2-tagged proteins for 16 hours in the presence of 100 nM Taxol and 50 μ M of Biotin. BioID2 biochemical purifications were then performed with streptavidin beads to capture biotinylated proteins. Diagnostic samples of each purification (low speed supernatant (LS), high speed supernatant (HS), eluate, and unbound) were resolved by western blotting and the blots were probed with streptavidin to visualize the biotinylated eluates of each BioID2-tagged SAC protein purification. (B) BioID2-Myc (empty vector, EV), BioID2-Myc-MAD1L1, or BioID2-Myc-MAD2L1 inducible HeLa stable cell lines were induced with Dox and treated with 100 nM Taxol to arrest cells in mitosis. Mitotic cell lysates were then used for reciprocal Myc immunoprecipitations and subjected to immunoblot analysis with the indicated antibodies. Note that endogenous MAD1L1 immunoprecipitates with BioID2-Myc-MAD2L1 and endogenous MAD2L1 immunoprecipitates with BioID2-Myc-MAD1L1. Asterisks denote endogenous MAD2L1 and BioID2-Myc-MAD2L1.

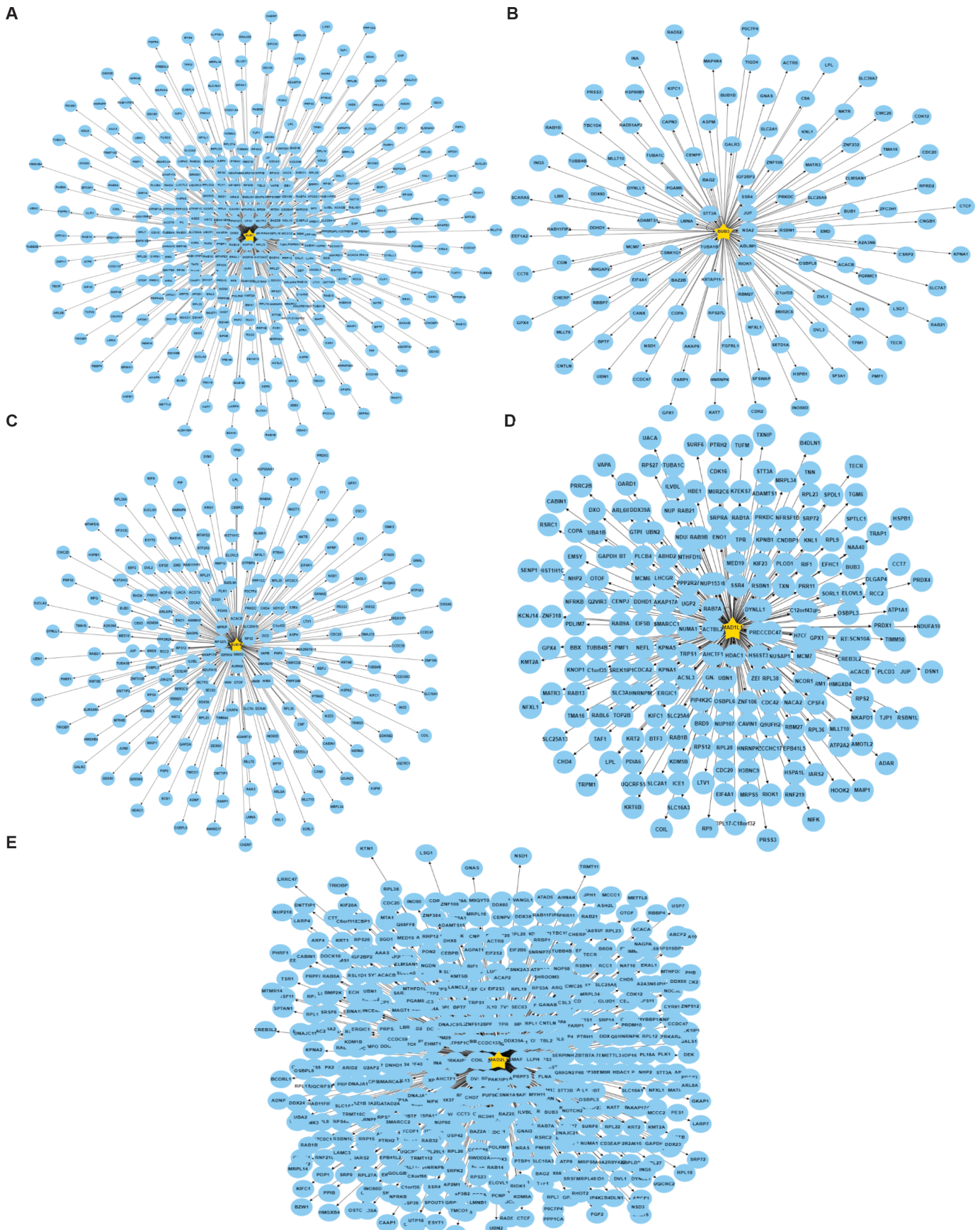


Figure S5. Core SAC protein proximity association maps. (A-E) Protein proximity association maps for each of the core SAC proteins BUB1 (A), BUB3 (B), BUBR1 (C), MAD1L1 (D), and MAD2L1 (E). See Table S4 for a list of proteins in each map. The maps were visualized using R-CytoscapeJS.

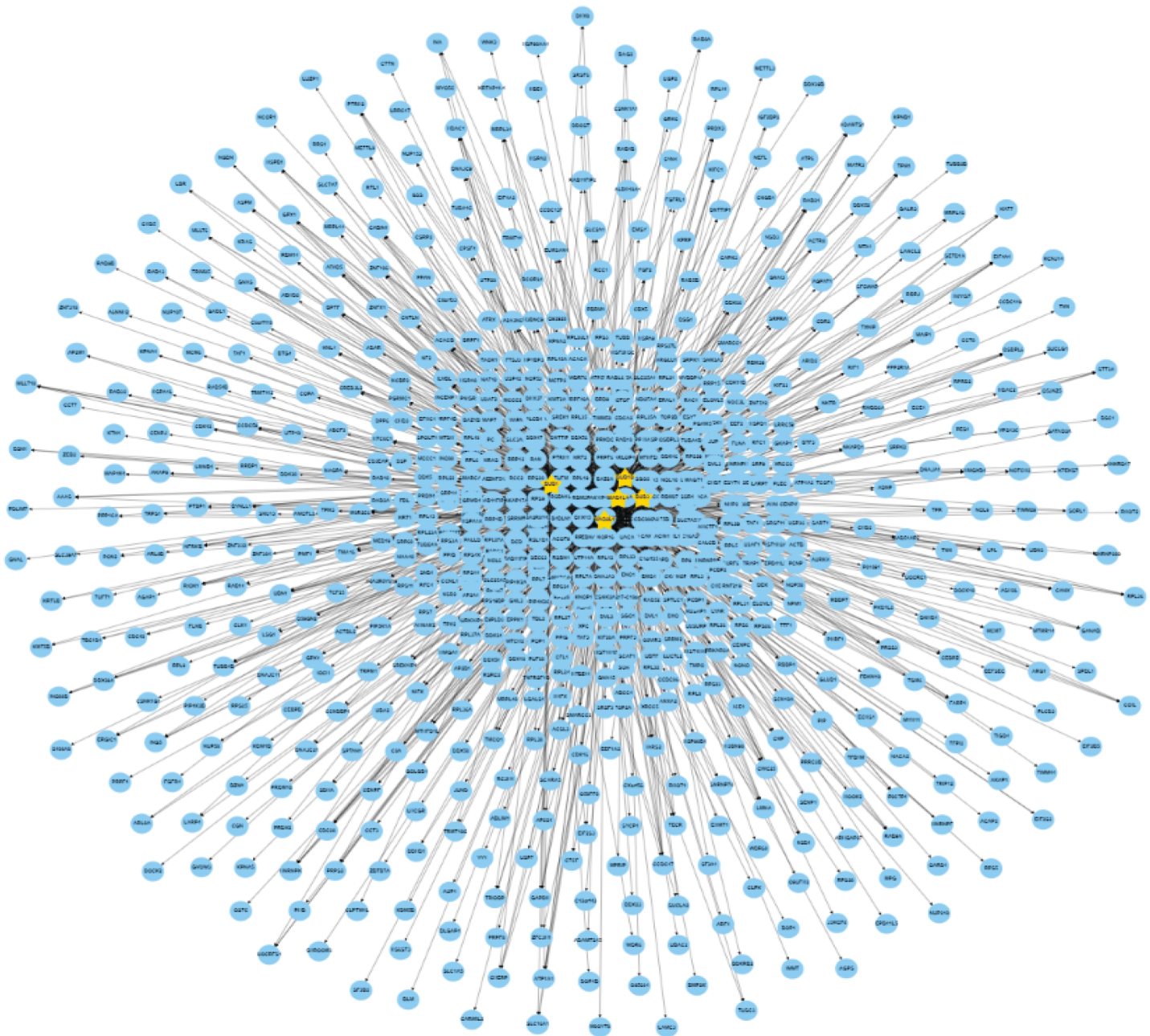


Figure S6. Core SAC protein proximity network. The five core SAC protein (BUB1; BUB3; BUBR1; MAD1L1; MAD2L1) proximity association maps in Figure S5 were combined to generate the core SAC protein proximity association network. The network was visualized using RCytoscapeJS.

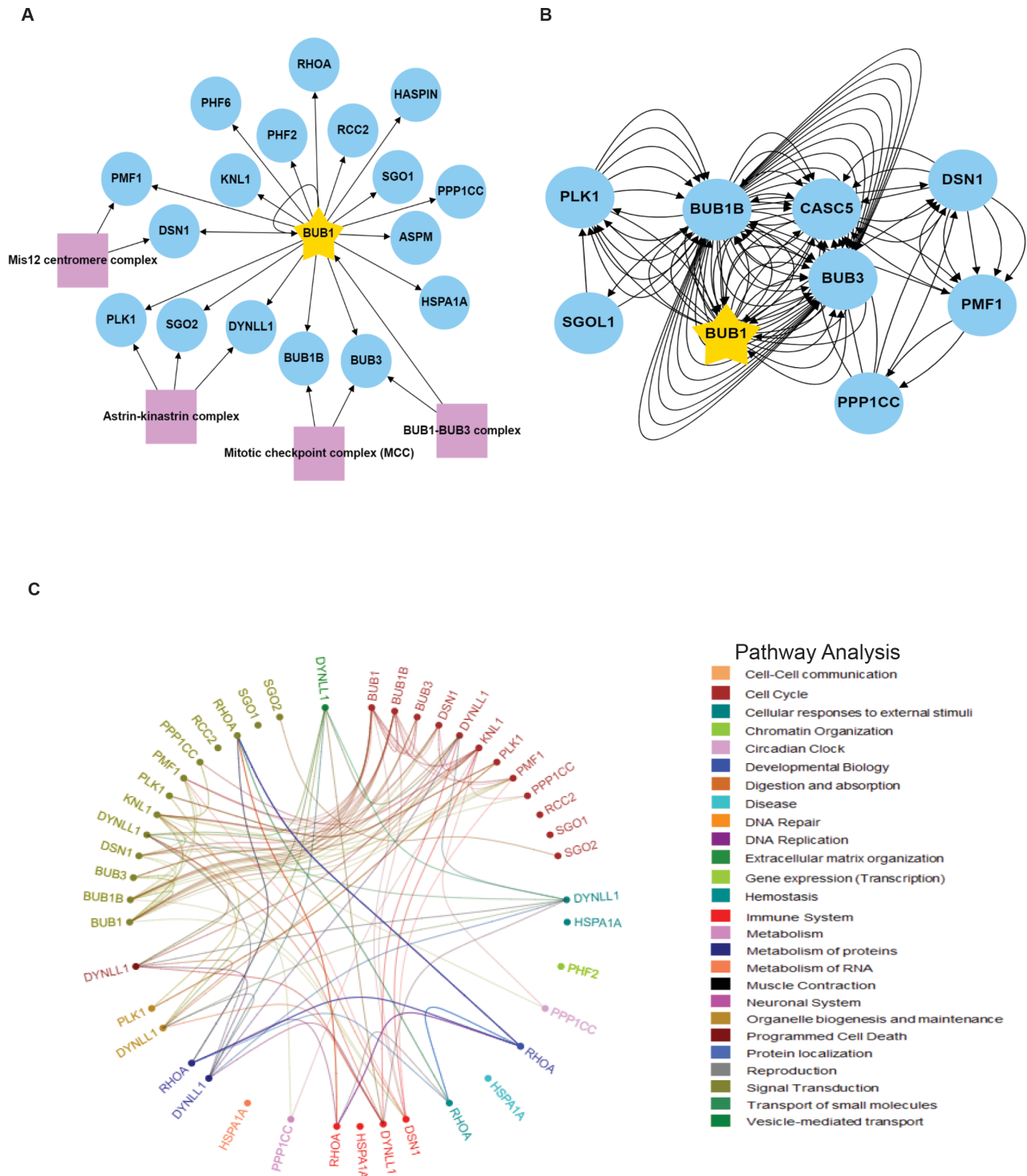


Figure S7. The BioID2-BUB1 proximity protein association map. (A) Generation of the BUB1 protein proximity association map using kinetochore related Gene Ontology annotations and CORUM complex annotation analyses. The map was visualized using RCytoscapeJS. Purple boxes highlight protein complexes known to assemble with BUB1 as annotated by the CORUM database. Arrows indicate the direction of the detected interactions. (B) The BUB1 proximity association map was analyzed with BioGRID to reveal previously verified protein associations. Each arrow indicates an experimentally annotated interaction curated in the BioGRID database. Direction of arrows indicate an annotated interaction from a bait protein to the prey. (C) Reactome pathway analysis of the BUB1 proximity association map. The Reactome circular interaction plot depicts the associations between the identified proteins within the BUB1 proximity association map and the corresponding pathways in which they function. Legend presents color-coded pathways that correspond to the circular interaction plot.

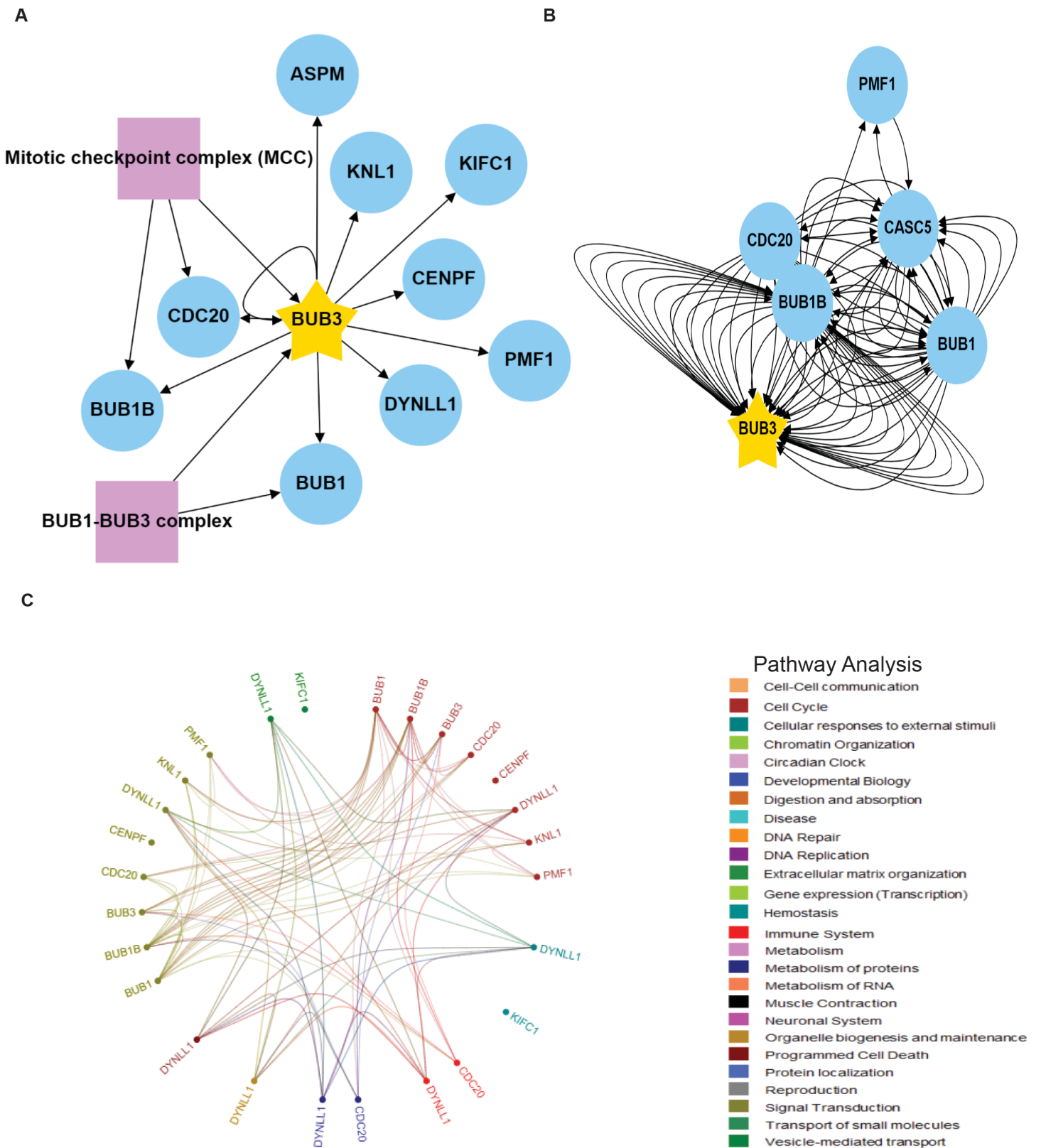


Figure S8. The BioID2-BUB3 proximity protein association map. A, B, and C are as described in Figure S7.

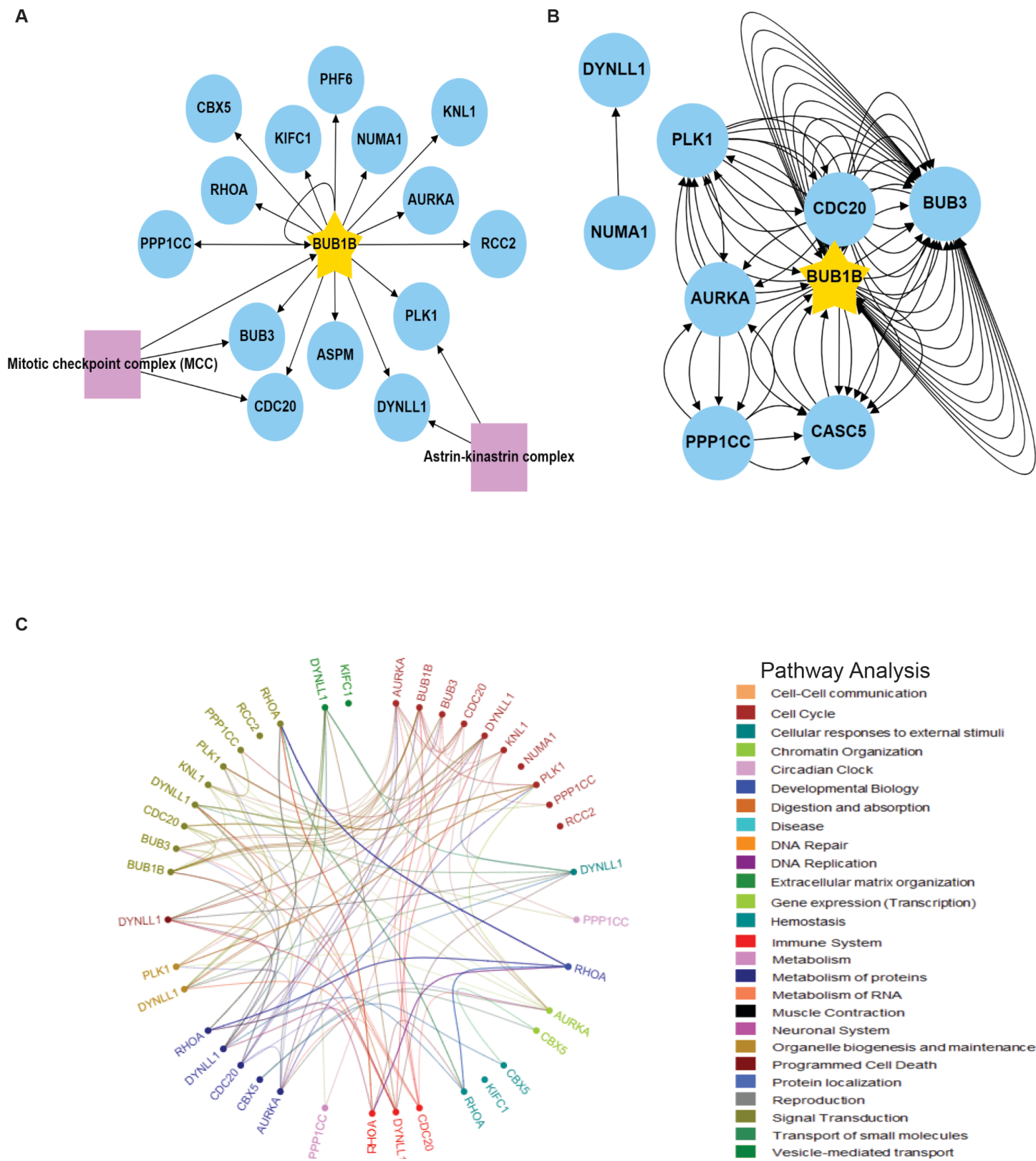


Figure S9. The BioID2-BUBR1/BUB1B proximity protein association map. A, B, and C are as described in Figure S7.

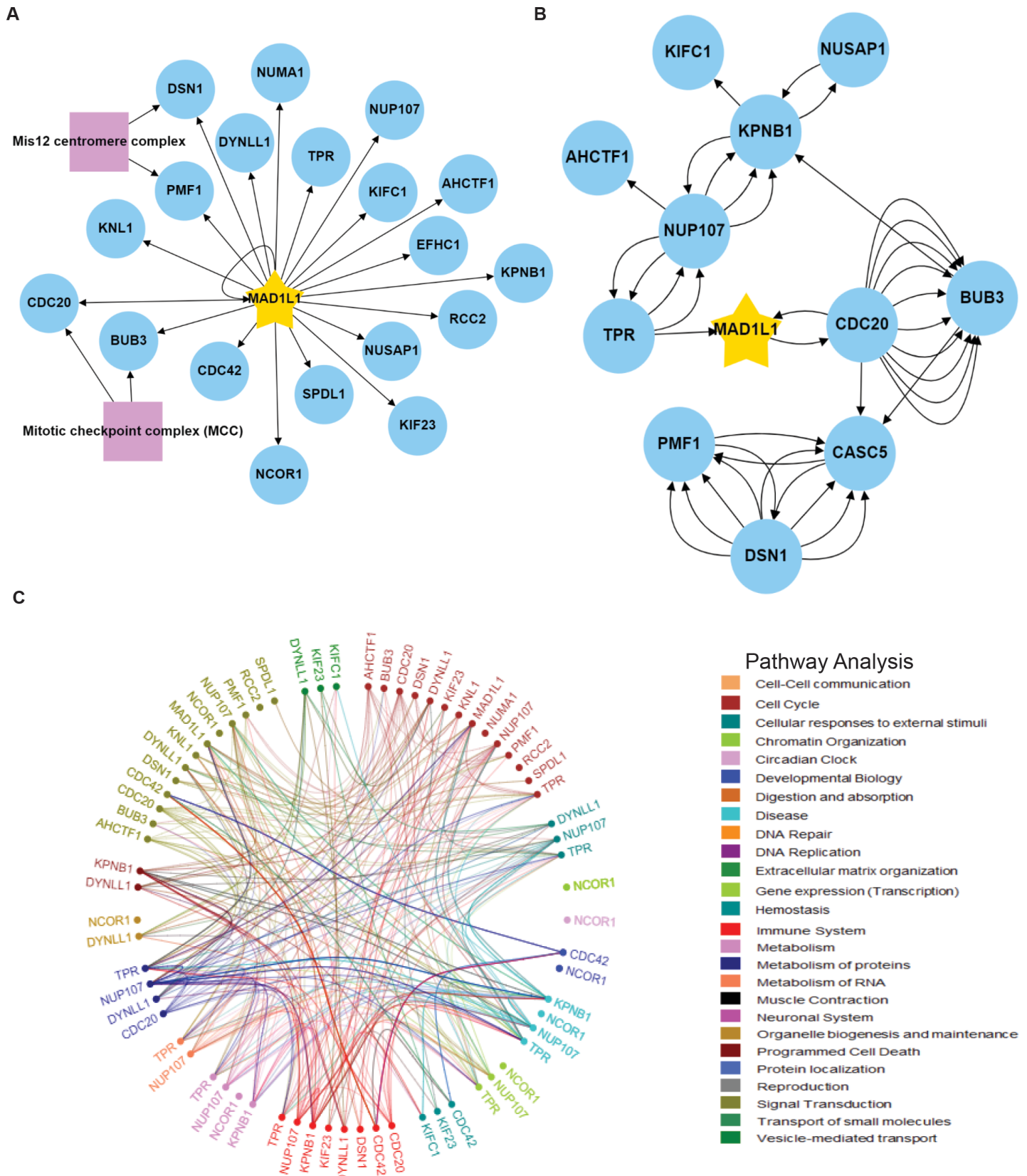


Figure S10. The BioID2-MAD1L1 proximity protein association map. A, B, and C are as described in Figure S7.

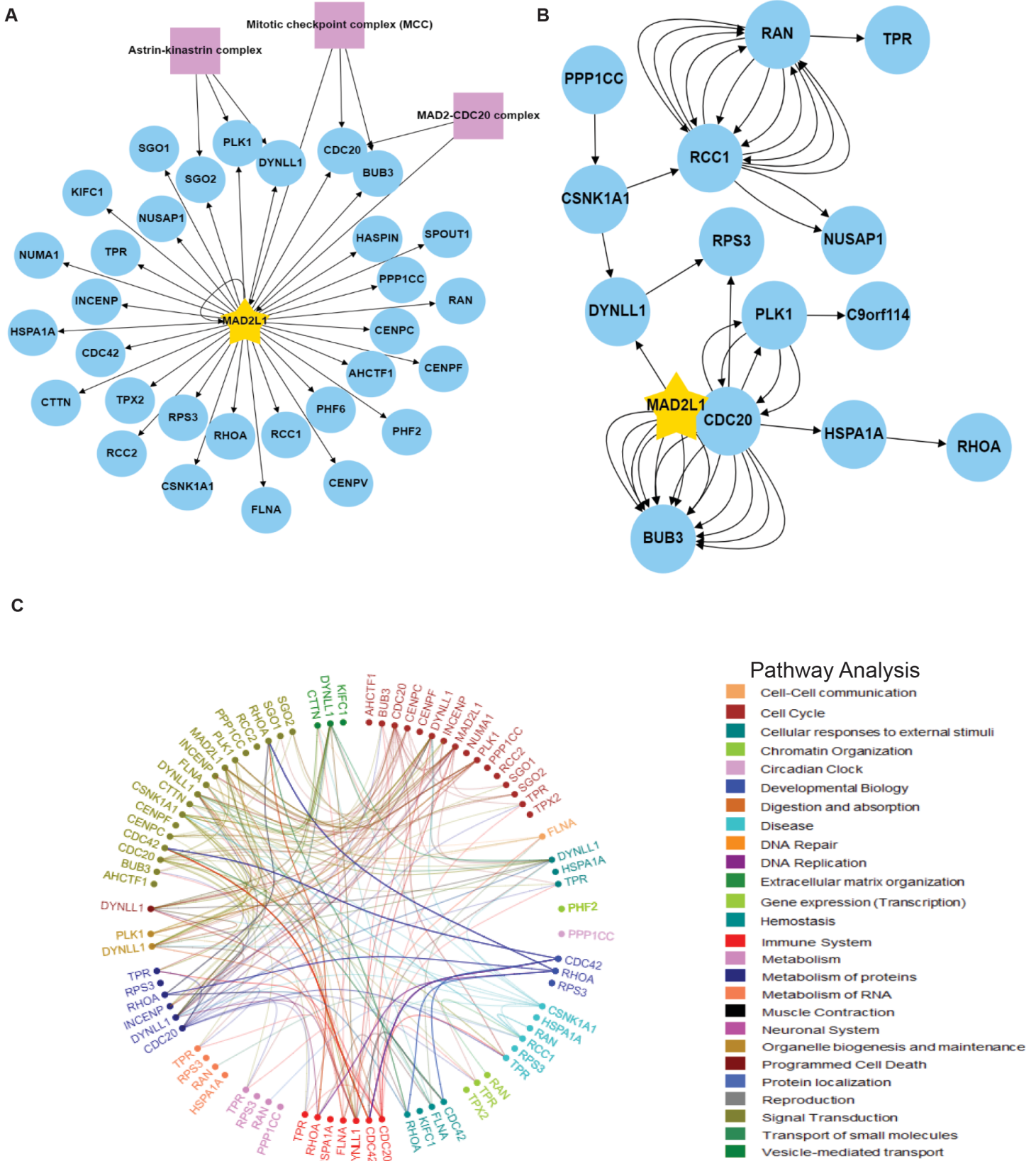
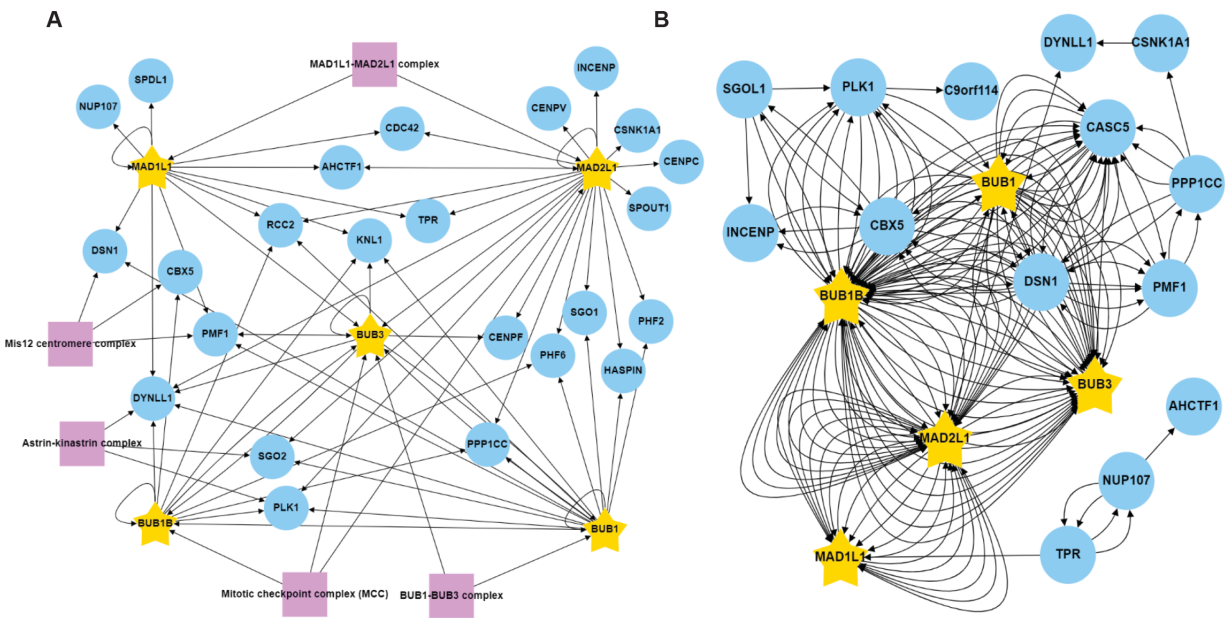


Figure S11. The BioID2-MAD2L1 proximity protein association map. A, B, and C are as described in Figure S7.



C

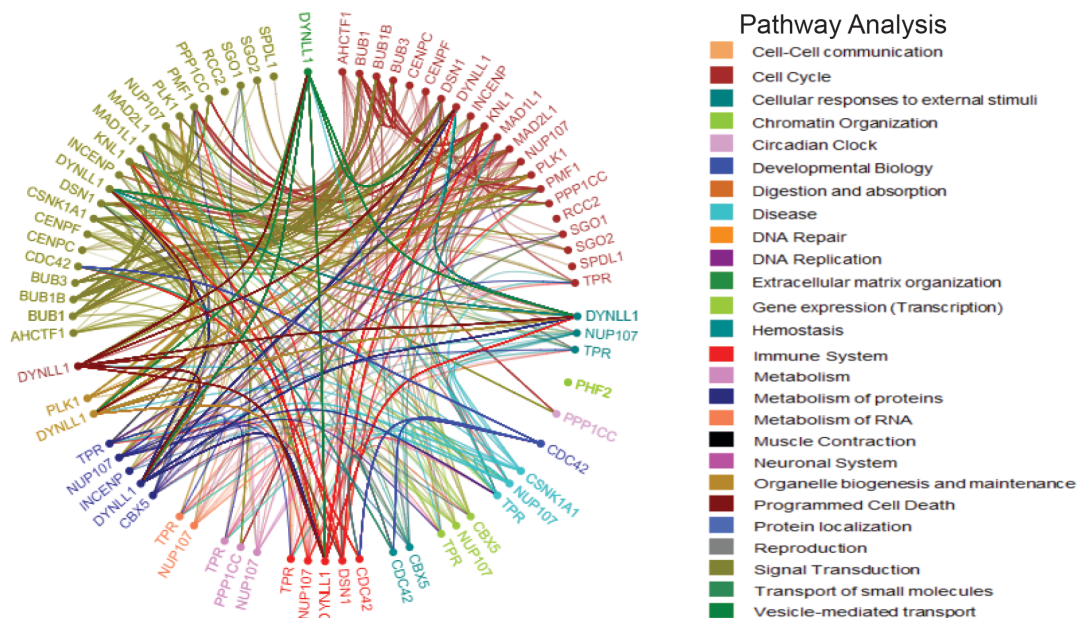


Figure S12. The core SAC protein proximity association network using kinetochore Gene Ontology annotations. (A) Generation of the core SAC protein (BUB1; BUB3; BUBR1; MAD1L1; MAD2L1) proximity association network using kinetochore related Gene Ontology annotations and CORUM complex annotation analyses. The map was visualized using RCytoscapeJS. Purple boxes highlight protein complexes known to assembly with core SAC proteins as annotated by the CORUM database. Arrows indicate the direction of the detected interactions. (B) The core SAC protein kinetochore proximity association network was analyzed with BioGRID to reveal previously verified protein associations. Each arrow indicates an experimentally annotated interaction curated in the BioGRID database. Direction of arrows indicate an annotated interaction from a bait protein to the prey. (C) Reactome pathway analysis of the core SAC protein kinetochore proximity association network. The Reactome circular interaction plot depicts the associations between the identified proteins within the core SAC protein proximity association network and the corresponding pathways in which they function. Legend presents color-coded pathways that correspond to the circular interaction plot.

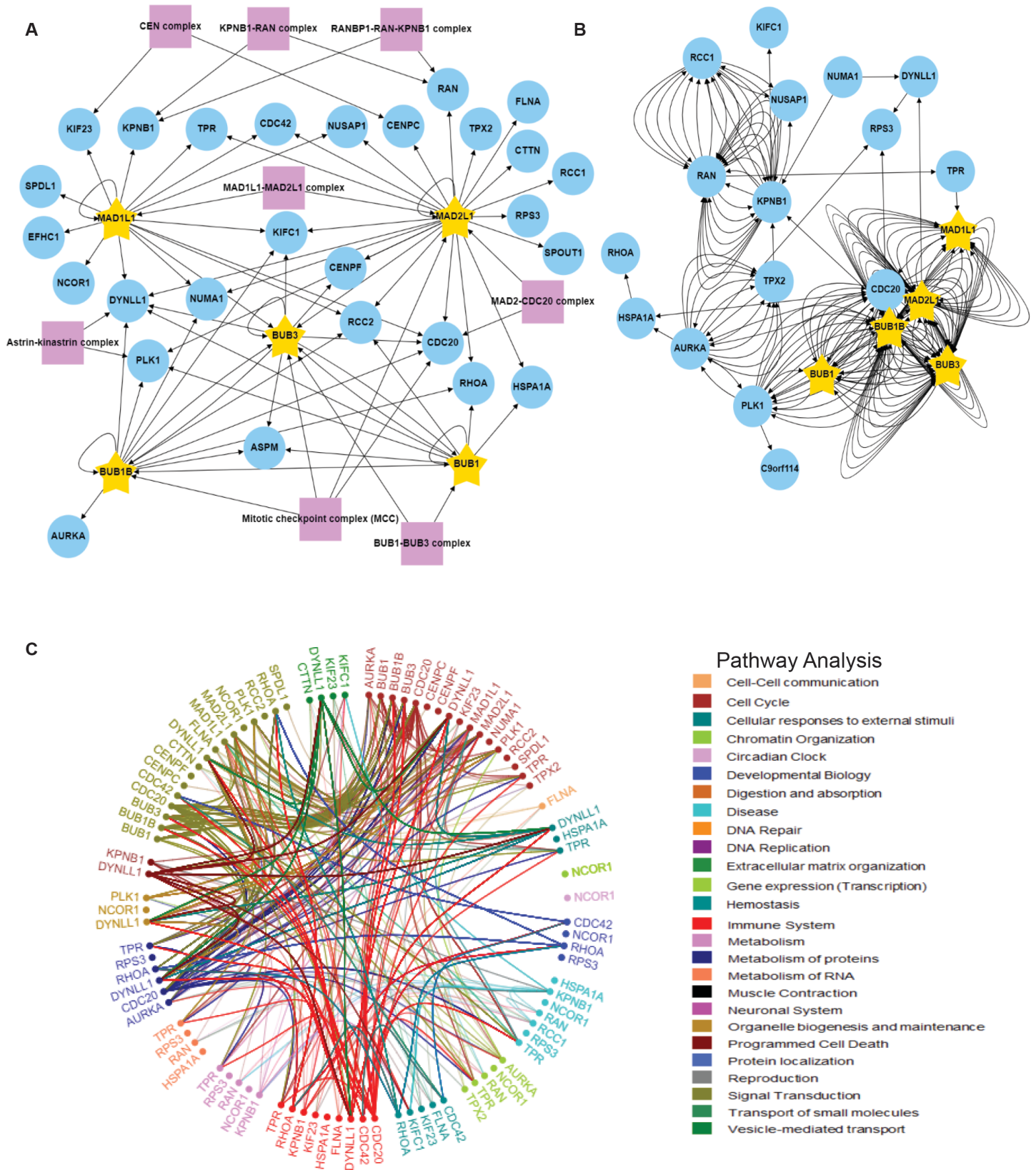


Figure S13. The core SAC protein proximity association network using mitotic spindle related Gene Ontology annotations. A, B, and C are as described in Figure S12, except that the mitotic spindle related Gene Ontology annotations were applied to the analysis.

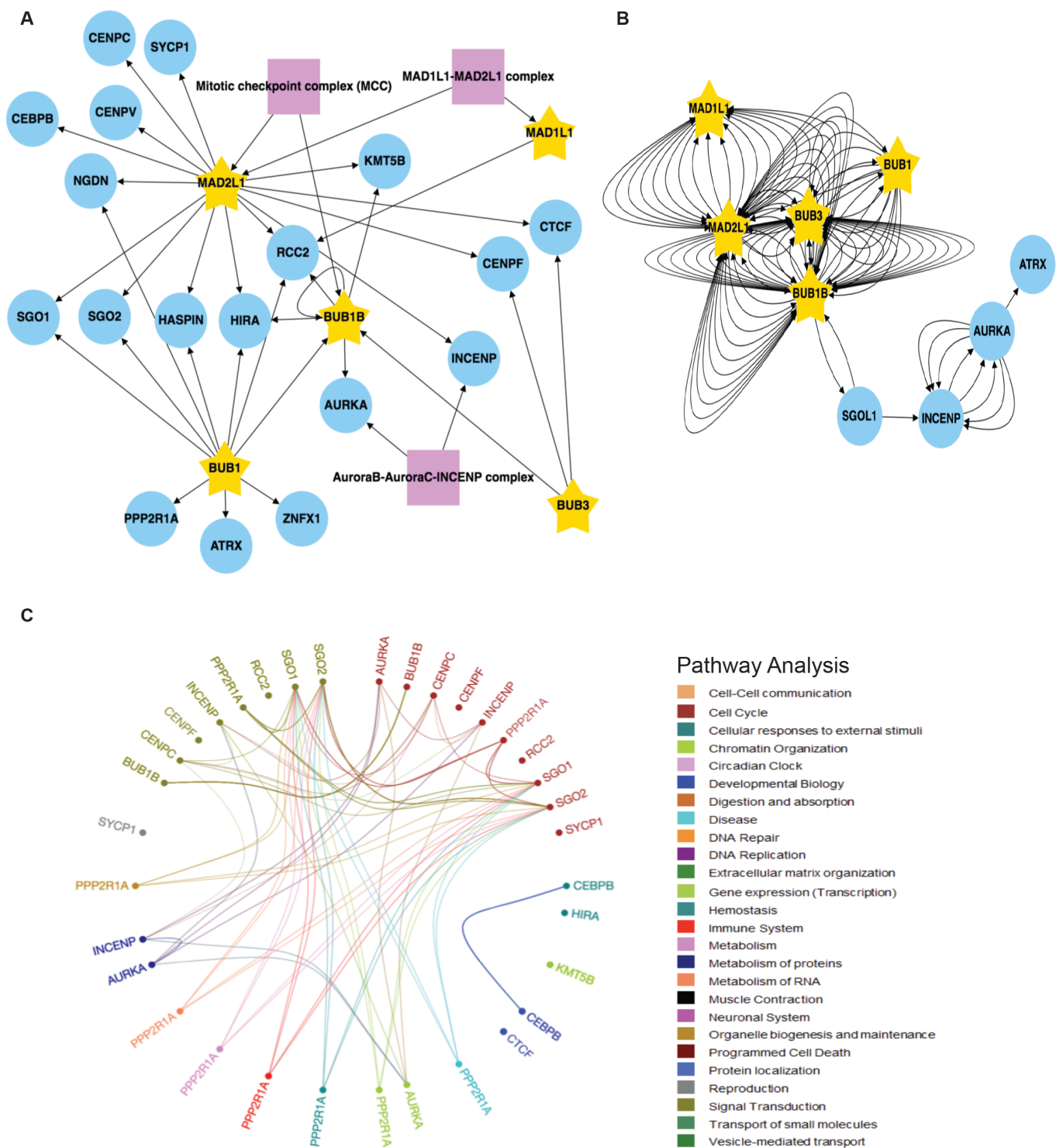


Figure S14. The core SAC protein proximity association network using centromere related Gene Ontology annotations. A, B, and C are as described in Figure S12, except that the centromere related Gene Ontology annotations were applied to the analysis.

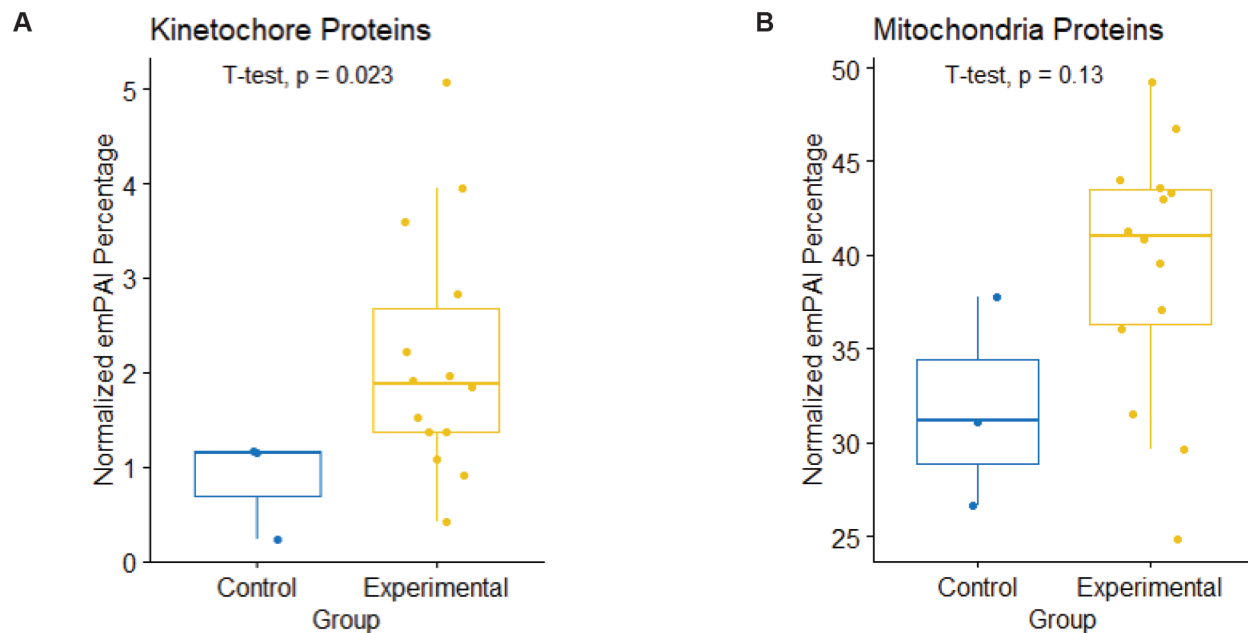


Figure S15. Kinetochores protein enrichment analysis. (A) Kinetochores related GO terms (see Table S8 for GO terms) were used to identify kinetochores related proteins and the normalized emPAI scores were used as a quantitative measure of the amount of each protein in each purification. Comparing the amount of kinetochores related proteins to the sum of normalized emPAI scores of all proteins in each purification resulted in a normalized emPAI percentage. This percentage was representative of the amount of kinetochores proteins compared to the total amount of proteins identified in the purification. This resulted in a range of 0.5- 5% kinetochores proteins in each experimental purification. These percentages were then compared to control purifications to determine whether there was an enrichment of kinetochores related proteins in the experimental purifications. After analyzing the control purifications as described above for the experimental, we performed a student's t-test to determine if there was a significant difference. The results indicate that there are lower amounts, 0.2-1%, of kinetochores related proteins in the control purifications, and that the difference is significant (p value = 0.023). (B) The same analysis in (A) was performed with Mitochondria related proteins. The results showed no significant difference between the percentage of mitochondria related proteins in control purifications compared to experimental purifications.

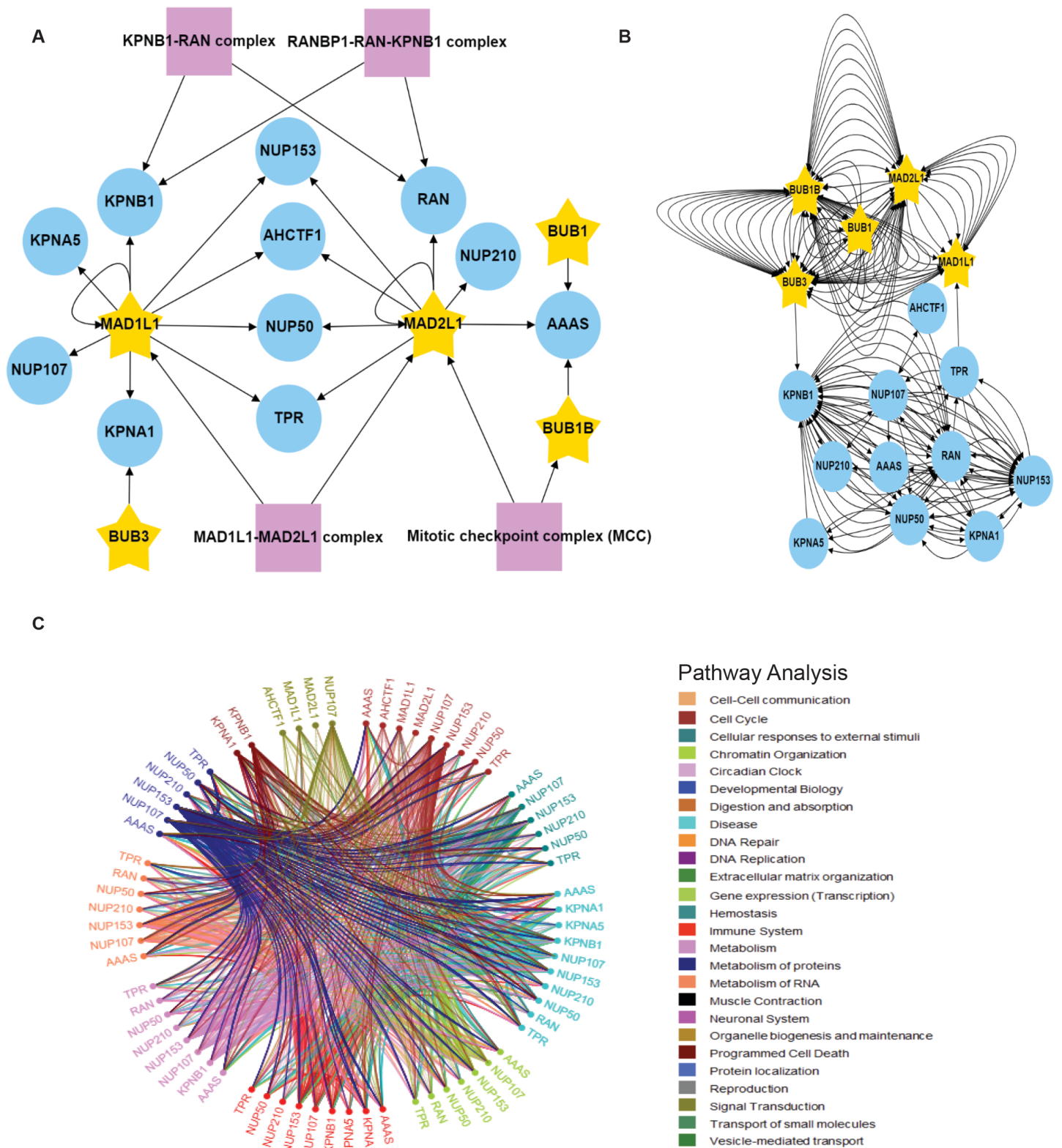


Figure S16. The core SAC protein proximity association network using nuclear pore related Gene Ontology annotations. A, B, and C are as described in Figure S12, except that nuclear pore related Gene Ontology annotations were applied to the analysis.

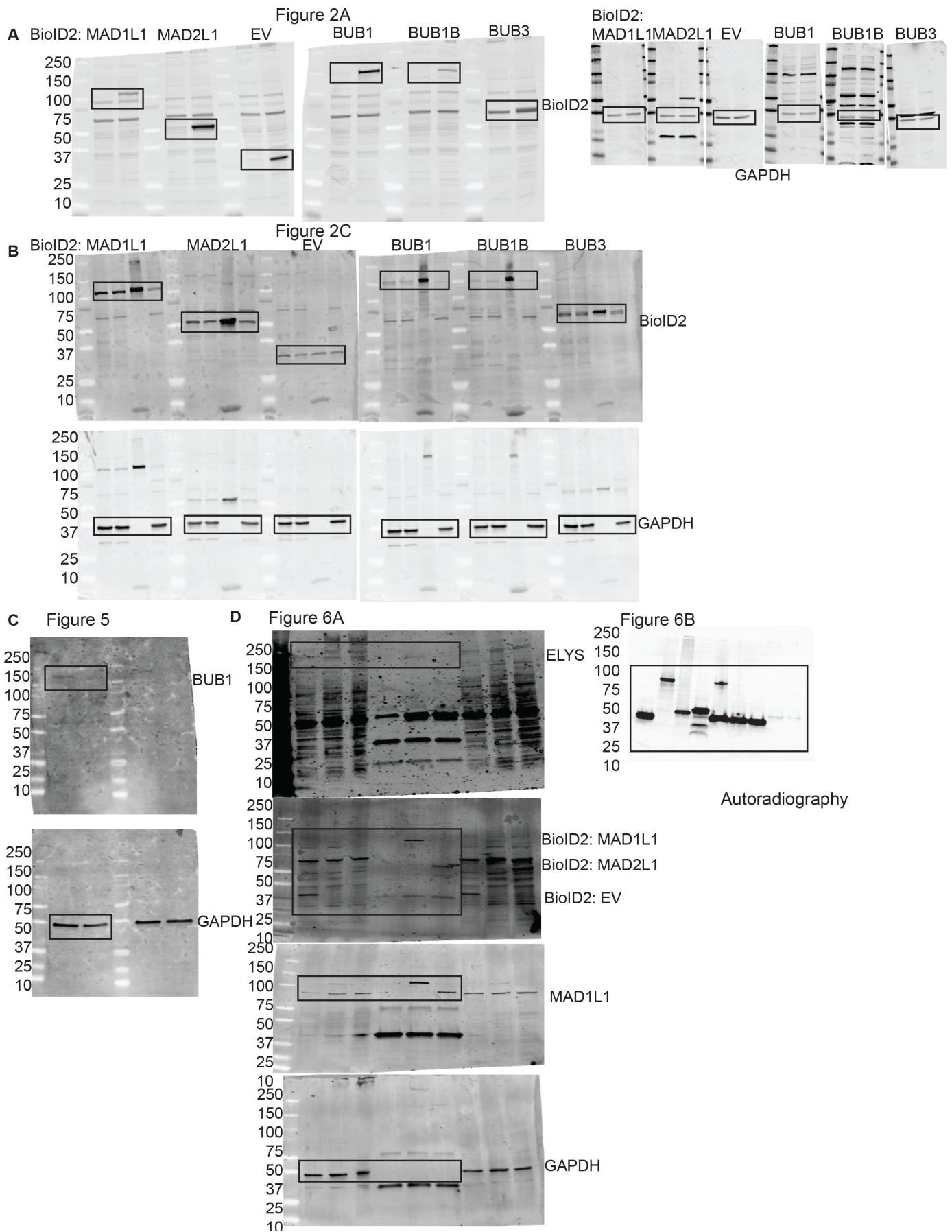
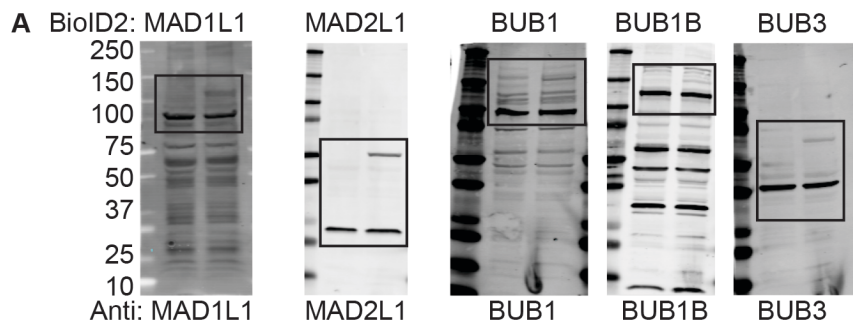


Figure S17. Uncropped immunoblots and autoradiography blots for all figures.

Supplemental Figure 3A



Supplemental Figure 4B

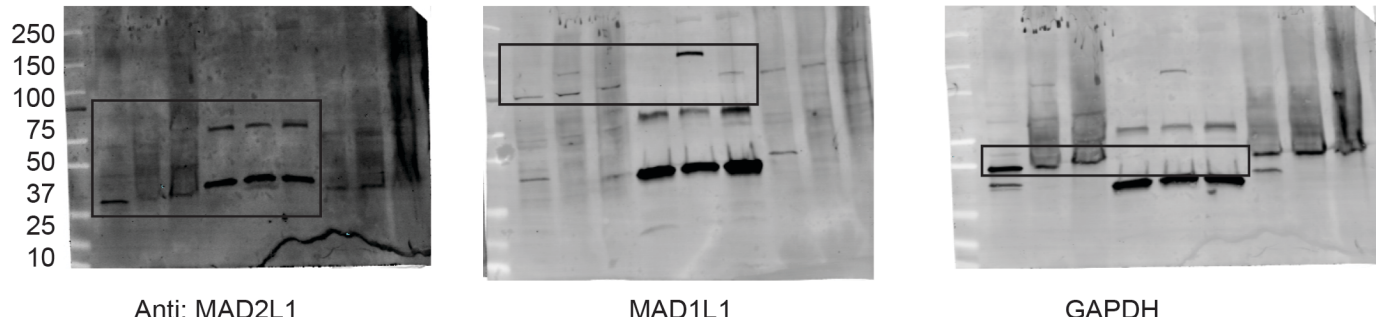


Figure S18. Uncropped immunoblots for all supplemental figures.



DUSP7 regulates the activity of ERK2 to promote proper chromosome alignment during cell division

Received for publication, November 10, 2020, and in revised form, April 10, 2021. Published, Papers in Press, April 16, 2021, <https://doi.org/10.1016/j.jbc.2021.100676>

Xiao Guo¹, Ivan Ramirez¹, Yenni A. Garcia¹, Erick F. Velasquez¹, Ankur A. Gholkar¹, Whitaker Cohn², Julian P. Whitelegge^{2,3,4} , Bobby Tofig⁵, Robert Damoiseaux^{5,6}, and Jorge Z. Torres^{1,3,4,*} 

From the ¹Department of Chemistry and Biochemistry, ²Pasarow Mass Spectrometry Laboratory, The Jane and Terry Semel Institute for Neuroscience and Human Behavior, David Geffen School of Medicine, ³Molecular Biology Institute, ⁴Jonsson Comprehensive Cancer Center, ⁵California NanoSystems Institute, and ⁶Department of Molecular and Medical Pharmacology, University of California, Los Angeles, California, USA

Edited by Enrique De La Cruz

Human cell division is a highly regulated process that relies on the accurate capture and movement of chromosomes to the metaphase plate. Errors in the fidelity of chromosome congression and alignment can lead to improper chromosome segregation, which is correlated with aneuploidy and tumorigenesis. These processes are known to be regulated by extracellular signal-regulated kinase 2 (ERK2) in other species, but the role of ERK2 in mitosis in mammals remains unclear. Here, we have identified the dual-specificity phosphatase 7 (DUSP7), known to display selectivity for ERK2, as important in regulating chromosome alignment. During mitosis, DUSP7 bound to ERK2 and regulated the abundance of active phospho-ERK2 through its phosphatase activity. Overexpression of DUSP7, but not catalytically inactive mutants, led to a decrease in the levels of phospho-ERK2 and mitotic chromosome misalignment, while knockdown of DUSP7 also led to defective chromosome congression that resulted in a prolonged mitosis. Consistently, knockdown or chemical inhibition of ERK2 or chemical inhibition of the MEK kinase that phosphorylates ERK2 led to chromosome alignment defects. Our results support a model wherein MEK-mediated phosphorylation and DUSP7-mediated dephosphorylation regulate the levels of active phospho-ERK2 to promote proper cell division.

Critical to the fidelity of cell division is the accurate movement and alignment of chromosomes at the metaphase plate and their segregation during anaphase. Errors in these processes are linked to human developmental disorders and tumorigenesis (1). Previous research has underscored the importance of protein phosphorylation as a molecular switch to regulate the activity of cell division enzymes (2, 3). This is highlighted by the growing list of essential mitotic kinases and their substrates that carry out functions related to bipolar spindle assembly, kinetochore-microtubule attachment, chromosome congression, and chromosome segregation (4–6). Beyond well-established mitotic kinases, less studied phospho signaling pathways have been implicated in cell division including the

Wnt, mTOR, and MAPK/ERK pathways, among which MAPK/ERK is phosphorylated by MEKs (mitogen-activated protein kinase or extracellular signal-regulated kinase kinase) to regulate downstream transcription factors (7–9). In *Xenopus laevis* ERK2 (extracellular signal-regulated kinase 2) is critical for the spindle assembly checkpoint (10–12). In mammalian cells ERK1/2 activity is necessary for the G1/S transition and early G2 events for timely entry into mitosis (13, 14). However, whether human ERK2 is active in mitosis and what roles it plays in human somatic cell division remains ambiguous.

Our RNAi screen for novel factors important for cell division identified the dual-specificity phosphatase 7 (DUSP7/MKP-X). DUSP7, DUSP6/MKP-3, and DUSP9/MKP-4 are members of the cytoplasmic ERK specific mitogen-activated protein kinase phosphatases (MKPs) subfamily that share similar amino acid sequences, subcellular localizations, and substrate preferences (15–17). DUSPs can dephosphorylate both tyrosine and serine/threonine residues and are important modulators of signaling pathways that regulate cellular processes such as proliferation and apoptosis (16, 17). DUSP7 exhibits selectivity toward ERK1/2 (18–20) and is a regulator of oocyte meiosis (21–23). DUSP7 contains an N terminal noncatalytic Rhodanese-like domain and a C-terminal dual-phosphatase domain. A conserved Kinase Interaction Motif (KIM) in the noncatalytic domain is essential for the interaction between MKPs and ERK (19, 24). Two key amino acid residues within the conserved catalytic sequence (H/V)C(X₅)R(S/T) of the phosphatase domain, C331 and R337, are important for DUSP7's phosphatase activity (25, 26). However, in contrast to MKPs such as DUSP6 and DUSP9, little is known about the physiological functions of DUSP7.

Here, we have determined that MEK phosphorylation activity and DUSP7 phosphatase activity regulate the levels of active phospho-ERK2, which is important for the fidelity of chromosome alignment and segregation during cell division.

Results

DUSP7 interacts with ERK2 and regulates the levels of phospho-ERK2

To understand the role of DUSP7 during cell division, we began by defining the protein–protein interaction and protein

* For correspondence: Jorge Z. Torres, torres@chem.ucla.edu.

proximity networks of DUSP7 in mitotic cells. Localization and affinity purification (LAP = GFP-Tev-S-tag)-tagged and biotin identification 2 (BioID2)-tagged DUSP7 inducible HeLa stable cell lines were used to express LAP/BioID2-DUSP7 and biochemical purifications were analyzed by mass spectrometry. In-house R scripts were used to analyze the mass spectrometry data and protein interaction and proximity networks were visualized with RCytoscape JS (Fig. S1, A and B). Further, we applied Gene Ontology (GO) terms (mitotic spindle; kinetochore and chromosome segregation) (Fig. S1C) and CORUM complex annotation analyses to these networks (see [Experimental procedures](#) for details). These analyses determined that ERK2 (aka MAPK1) was also associating with DUSP7 in mitosis (Fig. 1, A and B). Next, we validated the DUSP7-ERK2 mitotic interaction by immunoprecipitation (IP) experiments using mitotic cell extracts from Taxol- or nocodazole-arrested LAP-DUSP7 stable cell lines (Fig. 1C).

The KIM domain was shown to be essential for the interaction of some DUSPs (DUSP1, 4, 6) with ERK2 (24, 27, 28), but it remained unknown which domain of DUSP7 bound to ERK2 and whether its KIM was required for binding to ERK2 or its ability to dephosphorylate ERK2. To better understand the DUSP7-ERK2 interaction, we generated DUSP7 KIM mutants R102A, R103A, and R102,103A double mutants (Fig. 1D and Fig. S2, A–C). IP experiments from mitotic cells transiently transfected with DUSP7 or DUSP7 KIM mutants showed that ERK2 IPed with DUSP7 but not DUSP7 KIM mutants (Fig. 1E). To further define the interaction domains of DUSP7 involved in ERK2 binding, we generated a series of LAP-DUSP7 stable cell lines expressing DUSP7 truncations (Fig. S3A). ERK2 failed to associate with DUSP7 truncations (Fig. S3B), likely due to DUSP7 destabilization. Next, we sought to determine the significance of the DUSP7-ERK2 interaction. Consistent with the abolished interaction between DUSP7 KIM mutants and ERK2 (Fig. 1E), DUSP7-R103A and DUSP7-R102,103A double mutants showed a slightly reduced ability to dephosphorylate ERK2 in mitotic HeLa cells (Fig. 1F and Fig. S3, C–E). However, DUSP7-R102A could still dephosphorylate ERK2 (Fig. 1F); this phenomenon was also observed for conserved KIM mutations in DUSP6 (29). Similarly, IP experiments using *in vitro* expressed proteins or mitotic cell extracts from DUSP7 or DUSP7 catalytic dead mutant (C331A and R337A) (Fig. S2, D–G) cell lines showed that ERK2 IPed with DUSP7 and DUSP7-R337A but not DUSP7-C331A (Fig. 1, G and H, Fig. S3, F–H). While overexpression of DUSP7 led to the absence of phospho-ERK2, overexpressed DUSP7-R337A or DUSP7-C331A showed a reduced ability to dephosphorylate ERK2 in mitotic HeLa cells (Fig. 1I). Together, these results showed that DUSP7 was binding to ERK2 during mitosis and that the DUSP7 KIM was required for this interaction, while the DUSP7 catalytic sites (C331 and R337) within its phosphatase domain were important for regulating the levels of active phospho-ERK2.

Knockdown of DUSP7 leads to chromosome alignment and segregation defects

To understand the importance of DUSP7's function in regulating the levels of active phospho-ERK2 during cell

division, we first identified siRNAs capable of depleting DUSP7 levels by immunoblot analysis (Fig. 2A and Fig. S4A) and DUSP7 mRNA expression by RT-qPCR (Fig. S4, B and C). Knockdown of DUSP7 led to a failure to dephosphorylate ERK2 and an increase in phospho-ERK2 levels (Fig. 2A). We then analyzed the consequences of depleting DUSP7 during metaphase (Fig. 2B) and postmetaphase (Fig. 2E) with immunofluorescence (IF) microscopy. DUSP7 depletion led to an increased percentage of defective mitotic cells with chromosome misalignment (siDUSP7 = 44.6 ± 5.6 , $p < 0.05$ compared with siControl = 29.1 ± 2.9) (Fig. 2, C and D). These defective cells also showed defects in spindle organization including unfocused and multipolar spindles (Fig. 2C). The chromosome misalignments defects in siDUSP7 cells translated into an increase in the percentage of lagging chromosomes during anaphase (siDUSP7 = 24.9 ± 3.7 , $p < 0.05$ compared with siControl = 13.45 ± 3.1) (Fig. 2, F and G). The mitotic defects were rescued by an siRNA resistant DUSP7 (Fig. S2H) expressed at near endogenous levels but not DUSP7 catalytic dead mutants (Fig. S4, D–F). Similar results were observed and statistically analyzed in U2OS cells (Fig. S4, G–I) and HCT116 cells (Fig. S4, J–L).

Next, we analyzed whether DUSP7 depletion could affect the timing of cell division by live-cell time-lapse microscopy in HCT116 GFP-H2B cells (Fig. 2H). This analysis showed that depletion of DUSP7 led to a marked increase in the time from chromosome condensation to chromosome segregation (siDUSP7 = 54.0 ± 38.3 min, $p < 0.01$ compared with siControl = 38.0 ± 19.1 min) (Fig. 2, I–K; [Movies S1–S4](#)). Together, these results showed that depletion of DUSP7 led to a slowed mitosis where cells failed to properly align and segregate chromosomes.

Downregulation of ERK2 leads to chromosome alignment defects

Next, we sought to determine if ERK2 was important for human cell division. First, we depleted endogenous ERK2 by RNAi (Fig. 3A) and analyzed the consequences during cell division (Fig. 3B) with IF microscopy. Depletion of ERK2 led to an increased number of cells with defects in chromosome alignment during metaphase (siERK2 = 52.1 ± 2.8 , $p < 0.01$ compared with siControl = 30.6 ± 3.5) (Fig. 3, C and D), which was consistent in U2OS cells (Fig. S5, A and B) and HCT116 cells (Fig. S5, C and D). ERK2 depletion also led to an increase in interphase cells that were multinucleated or contained micronuclei (siERK2 = 22.6 ± 6.2 , $p < 0.01$ compared with siControl = 3.8 ± 0.8) (Fig. S5, E and F).

Since phospho-ERK2 levels were lower in mitosis than in G1/S phase (Fig. S5, G–J), we asked if ERK2 phosphorylation or ERK2 kinase activity was important for cell division. HeLa cells were treated with a MEK inhibitor U0126 (30, 31) or the ERK2 ATP-competitive inhibitor FR180204 (32) and analyzed by western blotting and IF microscopy (Fig. 3E). Phospho-ERK2 levels decreased in U0126-treated cells, but were not affected in FR180204-treated cells (Fig. 3F). In comparison to the control DMSO treatment, cells treated with U0126 or FR180204 showed an increase in chromosome alignment

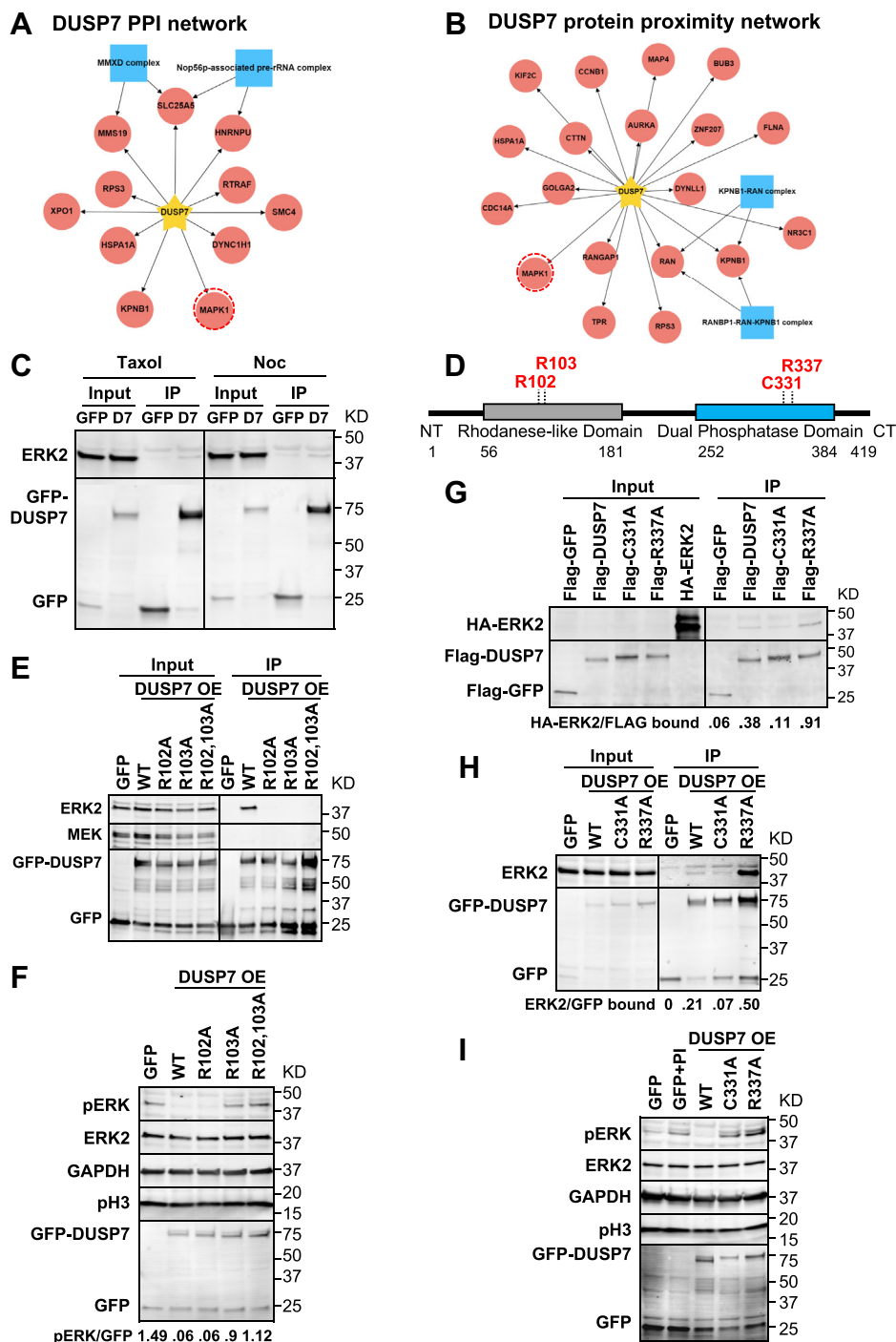


Figure 1. DUSP7 interacts with ERK2 and regulates the levels of phospho-ERK2. *A* and *B*, DUSP7 protein-protein interaction (PPI) (*A*) and protein proximity (*B*) networks generated using mitotic spindle GO annotations and CORUM complex annotation analyses. *Yellow stars* indicate the bait protein DUSP7; *red circles* indicate putative interactors; *blue squares* indicate protein complexes; *red dashed circles* highlight ERK2 (aka MAPK1). *C*, ERK2 immunoprecipitates (IPs) with DUSP7 (D7) in early (nocodazole (Noc) arrested cells) and mid (taxol (Tax) arrested cells) mitosis. *D*, schematic of DUSP7 domain structure and key sites. The number of amino acids are indicated for each domain. DUSP7 KIM (R102, R103) and catalytic sites (C331, R337) are in *red*. *E*, the DUSP7 KIM mediates the DUSP7-ERK2 mitotic interaction. *F*, the DUSP7 KIM is dispensable for its phosphatase activity. Ratios below immunoblots indicate normalized phospho-ERK2 levels. *G* and *H*, the DUSP7-ERK2 mitotic interaction is influenced by DUSP7's catalytic activity. In (*G*) HA-ERK2, Flag-DUSP7, Flag-C331A, Flag-R337A and Flag-GFP (negative control) were IVT expressed and incubated with anti-FLAG M2 magnetic beads in IP assays. In (*H*) LAP-only, LAP-DUSP7-WT, LAP-C331A, and LAP-R337A stable cell lines were induced before being harvested for S-tag pull downs. Ratios below immunoblots indicate relative protein-protein binding affinity. *I*, DUSP7 regulates mitotic phospho-ERK2 levels through its phosphatase activity. Phosphatase inhibitor (PI) in the second lane was added when lysing the cells. Numbers on the right side of immunoblots indicate molecular weights of proteins. All cell-based experiments and immunoprecipitations were carried out in HeLa cells.

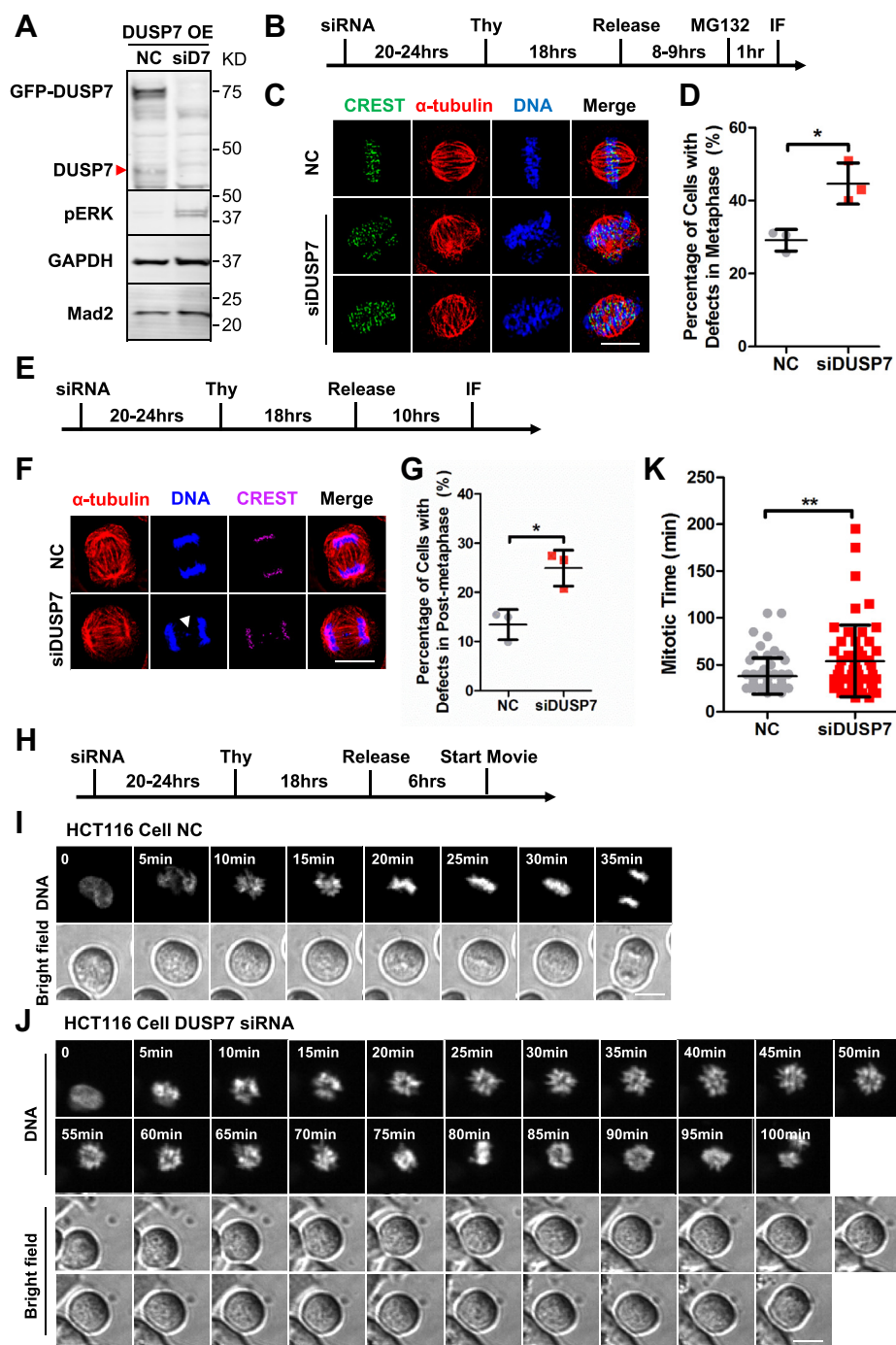


Figure 2. Knockdown of *DUSP7* leads to chromosome alignment and segregation defects. *A*, siRNA knockdown of endogenous and overexpressed *DUSP7*. Numbers on the right side of immunoblots indicate molecular weights of proteins. *Red arrow* indicates endogenous *DUSP7* band. *B*, schematic of IF microscopy experiment performed in (*C*). *C*, knockdown of *DUSP7* leads to chromosome misalignment in metaphase. HeLa cells were treated with negative control siRNA or si*DUSP7* before being fixed and costained with anti-CREST and anti- α -tubulin antibodies and the DNA dye Hoechst 33342. *D*, quantification of the percentage of cells with chromosome misalignment in metaphase (y-axis) for conditions shown in (*C*) (x-axis). *E*, schematic of IF microscopy experiment performed in (*F*). *F*, knockdown of *DUSP7* leads to an increase in lagging chromosomes in anaphase. HeLa cells were treated with negative control siRNA or si*DUSP7* before being fixed and costained with anti-CREST and anti- α -tubulin antibodies and the DNA dye Hoechst 33342. *White arrow* shows the lagging chromosome. *G*, quantification of the percentage of cells with lagging chromosome in anaphase (y-axis) for conditions shown in (*F*) (x-axis). *H*, schematic of live-cell time-lapse microscopy experiment performed in (*I*) and (*J*). *I* and *J*, knockdown of *DUSP7* leads to a slowed mitosis. Live-cell time-lapse microscopy of HCT116 GFP-H2B cells treated with negative control siRNA (*I*) and si*DUSP7* (*J*) undergoing cell division. *K*, quantification of the timing of mitosis from chromosome condensation to chromosome segregation (y-axis) for the conditions shown in (*I*) and (*J*) (x-axis). Scale bars: 10 μ m. * $p < 0.05$, ** $p < 0.01$ (unpaired two-tailed Student's *t*-test).

errors (U0126 = 48.7 ± 12.7 , $p < 0.05$ and FR180204 = 45.6 ± 6.5 , $p < 0.05$ compared with DMSO = 25.8 ± 3.9) (Fig. 3, *G* and *H*), which was consistent in U2OS cells (Fig. S5, *K* and *L*) and

HCT116 cells (Fig. S5, *M* and *N*). These results showed that inhibiting ERK2 phosphorylation, and thereby its activation, or ERK2's kinase activity led to chromosome alignment defects.

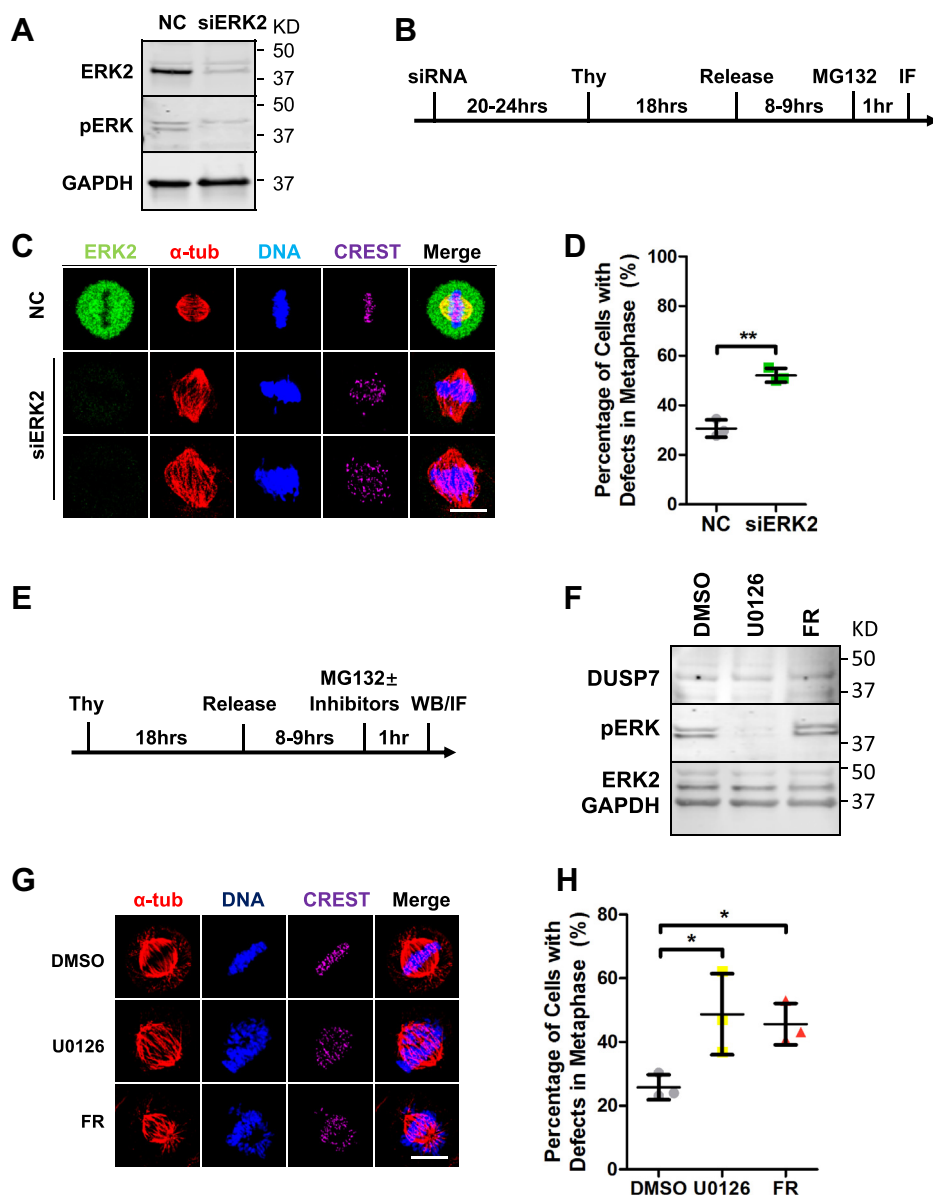


Figure 3. Downregulation of ERK2 leads to chromosome alignment defects. *A*, siRNA knockdown of ERK2. *B*, schematic of IF microscopy experiment performed in *C*. *C*, knockdown of ERK2 leads to chromosome misalignment in metaphase. HeLa cells were treated with negative control siRNA or siERK2 before being fixed and costained with anti-ERK2, anti-CREST and anti- α -tubulin antibodies and the DNA dye Hoechst 33342. *D*, quantification of the percentage of cells with chromosome misalignment in metaphase (y-axis) for conditions shown in *C* (x-axis). *E*, schematic of western blotting experiment performed in *F* and IF microscopy experiment performed in *G*. *F*, HeLa cells were treated with DMSO (as negative control), 50 μ M U0126, or 50 μ M FR180204 before being lysed and analyzed by immunoblot. *G*, inhibition of MEK kinase activity or ERK2 kinase activity leads to chromosome misalignment in metaphase. HeLa cells were treated with DMSO or the indicated inhibitors, fixed, and costained with anti-CREST and anti- α -tubulin antibodies and the DNA dye Hoechst 33342. *H*, quantification of the percentage of cells with chromosome misalignment in metaphase (y-axis) for the conditions shown in *G* (x-axis). Numbers on the right side of immunoblots indicate molecular weights of proteins. Scale bars: 10 μ m. * p < 0.05, ** p < 0.01 (unpaired two-tailed Student's *t*-test).

***DUSP7* promotes chromosome alignment in mitosis by regulating the activity of ERK2**

Since DUSP7 dephosphorylated ERK2 (Fig. 1E), we hypothesized that overexpression of DUSP7 would lead to similar chromosome alignment defects to those observed in cells treated with the MEK inhibitor U0126. To test this, we overexpressed GFP-tagged DUSP7 (validated to decrease phospho-ERK2 levels, Fig. 1, F and I) or the catalytic dead DUSP7-C331A or DUSP7-R337A mutants (showed minimal effects on phospho-ERK2 levels, Fig. 1I) and analyzed the cells by IF

microscopy (Fig. 4A). While DUSP7 overexpression led to a significant increase in chromosome alignment defects, overexpression of DUSP7-R337A or DUSP7-C331A did not (DUSP7 = 42.1 ± 6.9 , p < 0.05; DUSP7-C331A = 30.6 ± 3.5 , p = 0.3183; and DUSP7-R337A = 34.4 ± 5.7 , p = 0.1386; compared to the GFP control = 27.3 ± 3.5) (Fig. 4, B and C), which was consistent in U2OS cells (Fig. S6A) and HCT116 cells (Fig. S6B). These results showed that an overabundance of DUSP7 phosphatase activity led to chromosome alignment defects.

ACCELERATED COMMUNICATION: DUSP7 regulates ERK2 to promote cell division

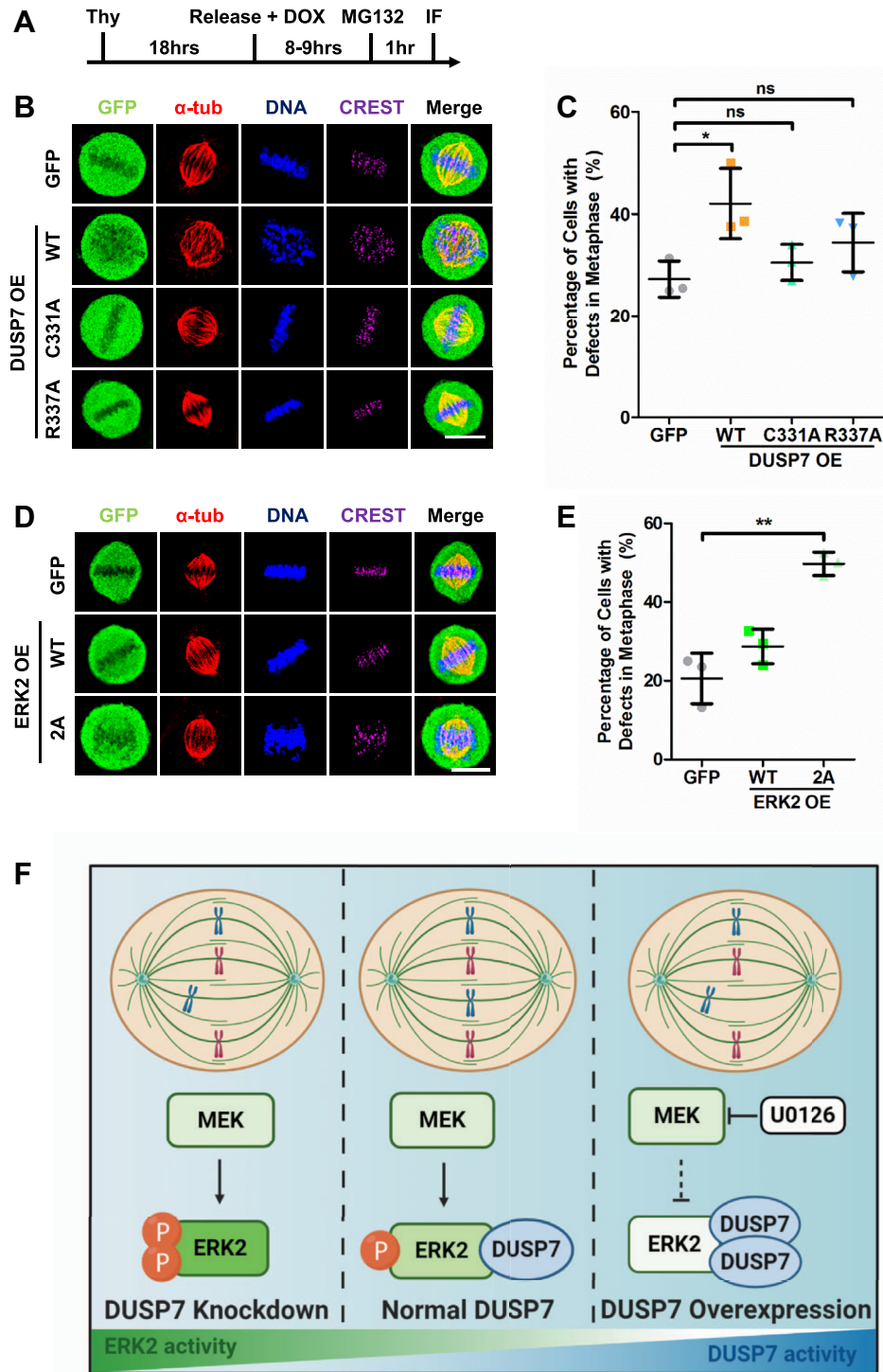


Figure 4. DUSP7 promotes chromosome alignment in mitosis by regulating the activity of ERK2. *A*, schematic of IF microscopy experiment performed in *(B)* and *(D)*. *B*, overexpression of DUSP7 wild type, but not catalytic dead mutants, leads to chromosome misalignment in metaphase. LAP-only, LAP-DUSP7-WT, LAP-C331A, and LAP-R337A HeLa stable cell lines were treated as described in *(A)* before being fixed and costained with anti-GFP, anti-CREST, and anti- α -tubulin antibodies and the DNA dye Hoechst 33342. *C*, quantification of the percentage of cells with chromosome misalignment in metaphase (y-axis) for conditions shown in *(B)* (x-axis). *D*, LAP-only, LAP-ERK2-WT, and LAP-ERK2-2A HeLa stable cell lines were treated as described in *(A)* before being fixed and costained with anti-GFP, anti-CREST, and anti- α -tubulin antibodies and the DNA dye Hoechst 33342. *E*, quantification of the percentage of cells with chromosome misalignment in metaphase (y-axis) for conditions shown in *(D)* (x-axis). *F*, model of how DUSP7 regulates the abundance of active phospho-ERK2 to ensure the fidelity of chromosome alignment. See main text for details. Scale bars: 10 μ m. * $p < 0.05$, ** $p < 0.01$, ns indicates not statistically significant (unpaired two-tailed Student's *t*-test).

To further understand the phospho-ERK2 equilibrium regulated by MEK and DUSP7 during cell division, we asked if the ERK2-DUSP7 interaction was dependent on ERK2 phosphorylation. IP experiments using cell extracts from a U0126-treated LAP-DUSP7 stable cell line showed that ERK2 bound to DUSP7 in the absence of MEK kinase activity (Fig. S6C). Since ERK2 is phosphorylated by MEK at T185 and Y187 (33, 34), we generated the nonphosphorylation mimetic mutant ERK2-2A (T185,Y187A) (Fig. S2, I and J) and analyzed its binding to DUSP7. *In vitro* binding experiments showed that both ERK2 and ERK2-2A bound to DUSP7 (Fig. S6D). Similar results were observed in IP experiments from HeLa cell extracts (Fig. S6E). Together, these results showed that DUSP7's binding to ERK2 did not require ERK2 to be phosphorylated. Instead, D318 within the ERK2 common docking (CD) domain (24) was responsible for its binding to DUSP7 (Fig. S6F). Next, we examined if phosphorylation of ERK2 at T185 and Y187 was critical for cell division by analyzing cells overexpressing GFP-tagged ERK2 or the nonphosphorylation mimetic mutant ERK2-2A (Fig. 4A). Compared with the overexpression of ERK2, overexpression of ERK2-2A led to a significant increase in cells with chromosome alignment defects in metaphase (ERK2 = 28.7 ± 4.4 , $p = 0.1438$; ERK2-2A = 49.7 ± 3.0 , $p < 0.01$; compared with GFP control = 20.6 ± 6.4) (Fig. 4, D and E), which was consistent in U2OS cells (Fig. S6A) and HCT116 cells (Fig. S6B). These data indicated that the proper amount of phospho-ERK2 in cells was critical for chromosome alignment and segregation during mitosis.

Discussion

This study revealed a novel function for DUSP7 in mitotic chromosome alignment and established the MAPK/ERK pathway as being important for cell division. Our data are consistent with a model where, during a normal mitosis, MEK's kinase activity phosphorylates ERK2 and DUSP7's phosphatase activity dephosphorylates ERK2 to establish an equilibrium of active phospho-ERK2 (Fig. 4F middle panel). This phospho-ERK2 equilibrium is critical for ensuring the fidelity of chromosome alignment and segregation. Perturbing the balance of active phospho-ERK2 through MEK inhibition (Fig. 4F right panel), DUSP7 depletion (Fig. 4F left panel) or overexpression (Fig. 4F right panel) leads to defects in chromosome alignment. Together, these results establish DUSP7 as an important mitotic phosphatase that regulates the abundance of active phospho-ERK2 to ensure the fidelity of chromosome alignment and segregation.

Interestingly, although the DUSP7 KIM mutant R102A did not bind ERK2, it could still dephosphorylate it (Fig. 1F, Fig. S3, C–E). This is consistent with previous DUSP6 observations, where the DUSP6 KIM mutant R64A did not bind ERK2 but was able to dephosphorylate it (29). Therefore, it is possible that these mutants are capable of transiently interacting with ERK2, but that the interaction is undetectable in IP experiments.

With the exception of ERK2, there is little known about the repertoire of DUSP7 substrates, regulators, and interactors. The GO enrichment analyses of the DUSP7 protein association network and DUSP7 proximity protein network indicate that

DUSP7 is likely to associate with numerous proteins that carry out important functions related to a broad array of cellular processes including apoptosis, regulation of transcription, and cell division (Table S1). Therefore, future studies aimed at understanding the importance of these interactions will further aid our understanding of DUSP7's function in cell division and beyond.

Experimental procedures

Cell culture

Table S2 lists all reagents and tools used in this study. HeLa cells (ATCC) were grown in DMEM/Ham's F-12 with L-Glutamine (Genesee Scientific), U2OS (ATCC) and HCT116 cells were grown in McCoy's 5A (Gibco), with 10% FBS and 5% CO₂ at 37 °C. Detailed experimental procedures for cell synchronization, cell transfection, and inhibitor treatments are in the Supporting information.

Generation of vectors and cell lines

DUSP7 and ERK2 mutants were generated by QuikChange Lightning Site-Directed Mutagenesis (Agilent). cDNAs of GFP, DUSP7, DUSP7 KIM mutants, DUSP7 catalytic dead mutants, ERK2, ERK2-2A, and DUSP7 truncations were cloned into pGLAP1, pGBioID2, pCS2-HA, or pCS2-Flag *via* Gateway LR Clonase reaction (35). pGLAP1-only/DUSP7/DUSP7-C331A/DUSP7-R337A/ERK2/ERK2-2A/DUSP7-truncations and pGBioID2-only/DUSP7 were used to generate Dox inducible HeLa Flp-In T-REx LAP-GFP/DUSP7-C331A/DUSP7-R337A/ERK2/ERK2-2A/DUSP7-truncations and HeLa Flp-In T-REx BioID2-only/DUSP7 stable cell lines as described previously (36, 37) (see Supporting information).

LAP/BioID2 purifications and LC-MS/MS analyses

LAP purifications from Taxol arrested LAP-tagged inducible stable cell lines were as previously described (36). For BioID2 purifications, biotinylated proteins were purified from Taxol-arrested BioID2-tagged inducible stable cell lines as described previously (38, 39). Mass spectrometry analysis was performed on a Thermo Q Exactive Plus Orbitrap as described previously (40). Protein–protein interaction information was integrated from the Biological General Repository for Interaction Datasets (BioGRID v. 3.5) (41). Protein-complex information was derived from the Comprehensive Resource of Mammalian Protein Complexes (CORUM v. 3.0) (42). Selected GO terms (Gene Ontology release June 2019) (43) were used to analyze the protein–protein interactions based on cellular mechanisms. Affinity-based and proximity-based networks were generated with RCytoscapeJS (44, 45). See Supporting information and Table S3–S6 for details on purifications, mass spectrometry, quantification of data, and protein interaction and proximity networks.

Immunoprecipitations, *in vitro* binding assays, and immunoblot analyses

Immunoprecipitations, *in vitro* binding assays, and immunoblot analyses were performed as described previously (46) with minor modifications detailed in the Supporting information.

ACCELERATED COMMUNICATION: *DUSP7* regulates ERK2 to promote cell division

Cell imaging

Fixed-cell and live-cell time-lapse microscopy was carried out as described previously (47), except that an ImageXpress XL imaging system (Molecular Devices) was used for live cell imaging. See [Supporting information](#) for details on imaging, quantification of data, and statistical analyses.

RT-qPCR

RNA from control or *DUSP7* siRNA transfected HeLa, U2OS, or HCT116 cells and *DUSP7* cell lines was isolated with Direct-zol RNA Miniprep Kits (Zymo Research) and reverse transcribed with UltraScript 2.0 cDNA Synthesis Kit (Genesee Scientific). qPCR was carried out with the synthesized cDNA, oligo(dT) primers, and qPCR BIO SyGreen Blue Mix Lo-ROX (Genesee Scientific) using a CFX Connect Real-Time PCR Detection System (Bio-Rad). qPCR data were analyzed with the Livak–Schmittgen method ($2^{-\Delta\Delta CT}$) (48).

Antibodies

See [Table S2](#) for a list of the antibodies used for immunoblotting and IF microscopy.

Data and code availability

Mass spectrometry data were deposited at the UCSD Center for Computational Mass Spectrometry MassIVE datasets <ftp://massive.ucsd.edu/MSV000085629/>. R scripts used to analyze and visualize LC-MS/MS results were deposited at GitHub <https://github.com/uclatorreslab/MassSpecAnalysis>. All remaining data are contained within this article.

Supporting information—This article contains [supporting information](#) (46, 49–51).

Acknowledgements—[Figure 4F](#) was created using [BioRender.com](#).

Author contributions—X. G., I. R., Y. A. G., E. F. V., A. A. G., W. C., J. P. W., B. T., R. D., and J. Z. T. performed the experiments, discussed the results, and wrote the paper.

Funding and additional information—Work was supported by NIH-NIGMS R35GM139539, R01GM117475 (J. Z. T.); USHHS T32CA009056 (Y. A. G.); a grant to the University of California, Los Angeles from the HHMI through the James H. Gilliam Fellowships for Advanced Study Program and an MBI Whitcome Fellowship (E. F. V.); NIH P30DK063491 (J. P. W.) and NCI P30CA016042 (R. D.). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Conflict of interest—The authors declare that they have no conflicts of interest with the contents of this article.

Abbreviations—The abbreviations used are: CT, C terminus; D7, *DUSP7*; DUSP, dual-specificity phosphatase; ERK, extracellular signal-regulated kinase; IF, immunofluorescence; IP, immunoprecipitation; KIM, kinase interaction motif; LAP, localization and

affinity purification; NC, negative control; NT, N terminus; OE, overexpression; PI, phosphatase inhibitor; WT, wild type; Thy, thymidine.


References

- Gordon, D. J., Resio, B., and Pellman, D. (2012) Causes and consequences of aneuploidy in cancer. *Nat. Rev. Genet.* **13**, 189–203
- Heim, A., Rymarczyk, B., and Mayer, T. U. (2017) Regulation of cell division. *Adv. Exp. Med. Biol.* **953**, 83–116
- Ong, J. Y., Bradley, M. C., and Torres, J. Z. (2020) Phospho-regulation of mitotic spindle assembly. *Cytoskeleton (Hoboken)* **77**, 558–578
- Saurin, A. T. (2018) Kinase and phosphatase cross-talk at the kinetochore. *Front. Cell Dev. Biol.* **6**, 62
- Magnaghi-Jaulin, L., Eot-Houllier, G., Gallaud, E., and Giet, R. (2019) Aurora A protein kinase: To the centrosome and beyond. *Biomolecules* **9**, 28
- Combes, G., Alharbi, I., Braga, L. G., and Elowe, S. (2017) Playing polo during mitosis: PLK1 takes the lead. *Oncogene* **36**, 4819–4827
- Bryja, V., Cervenka, I., and Cajanek, L. (2017) The connections of Wnt pathway components with cell cycle and centrosome: Side effects or a hidden logic? *Crit. Rev. Biochem. Mol. Biol.* **52**, 614–637
- Cuyas, E., Corominas-Faja, B., Joven, J., and Menendez, J. A. (2014) Cell cycle regulation by the nutrient-sensing mammalian target of rapamycin (mTOR) pathway. *Methods Mol. Biol.* **1170**, 113–144
- Eblen, S. T. (2018) Extracellular-regulated kinases: Signaling from Ras to ERK substrates to control biological outcomes. *Adv. Cancer Res.* **138**, 99–142
- Wang, X. M., Zhai, Y., and Ferrell, J. E., Jr. (1997) A role for mitogen-activated protein kinase in the spindle assembly checkpoint in XTC cells. *J. Cell Biol.* **137**, 433–443
- Minshull, J., Sun, H., Tonks, N. K., and Murray, A. W. (1994) A MAP kinase-dependent spindle assembly checkpoint in *Xenopus* egg extracts. *Cell* **79**, 475–486
- Takenaka, K., Gotoh, Y., and Nishida, E. (1997) MAP kinase is required for the spindle assembly checkpoint but is dispensable for the normal M phase entry and exit in *Xenopus* egg cell cycle extracts. *J. Cell Biol.* **136**, 1091–1097
- Shinohara, M., Mikhailov, A. V., Aguirre-Ghiso, J. A., and Rieder, C. L. (2006) Extracellular signal-regulated kinase 1/2 activity is not required in mammalian cells during late G2 for timely entry into or exit from mitosis. *Mol. Biol. Cell* **17**, 5227–5240
- Meloche, S., and Pouyssegur, J. (2007) The ERK1/2 mitogen-activated protein kinase pathway as a master regulator of the G1- to S-phase transition. *Oncogene* **26**, 3227–3239
- Seternes, O. M., Kidger, A. M., and Keyse, S. M. (2019) Dual-specificity MAP kinase phosphatases in health and disease. *Biochim. Biophys. Acta Mol. Cell Res.* **1866**, 124–143
- Kim, H. S., Fernandes, G., and Lee, C. W. (2016) Protein phosphatases involved in regulating mitosis: Facts and hypotheses. *Mol. Cells* **39**, 654–662
- Patterson, K. I., Brummer, T., O'Brien, P. M., and Daly, R. J. (2009) Dual-specificity phosphatases: Critical regulators with diverse cellular targets. *Biochem. J.* **418**, 475–489
- Keyse, S. M. (2008) Dual-specificity MAP kinase phosphatases (MKPs) and cancer. *Cancer Metastasis Rev.* **27**, 253–261
- Owens, D. M., and Keyse, S. M. (2007) Differential regulation of MAP kinase signalling by dual-specificity protein phosphatases. *Oncogene* **26**, 3203–3213
- Dowd, S., Sneddon, A. A., and Keyse, S. M. (1998) Isolation of the human genes encoding the pyst1 and Pyst2 phosphatases: Characterisation of Pyst2 as a cytosolic dual-specificity MAP kinase phosphatase and its catalytic activation by both MAP and SAP kinases. *J. Cell Sci.* **111**(Pt 22), 3389–3399
- Caunt, C. J., Armstrong, S. P., Rivers, C. A., Norman, M. R., and McArdle, C. A. (2008) Spatiotemporal regulation of ERK2 by dual specificity phosphatases. *J. Biol. Chem.* **283**, 26612–26623

22. Pfender, S., Kuznetsov, V., Pasternak, M., Tischer, T., Santhanam, B., and Schuh, M. (2015) Live imaging RNAi screen reveals genes essential for meiosis in mammalian oocytes. *Nature* **524**, 239–242
23. Tischer, T., and Schuh, M. (2016) The phosphatase *Dusp7* drives meiotic resumption and chromosome alignment in mouse oocytes. *Cell Rep.* **17**, 1426–1437
24. Tanoue, T., Adachi, M., Moriguchi, T., and Nishida, E. (2000) A conserved docking motif in MAP kinases common to substrates, activators and regulators. *Nat. Cell Biol.* **2**, 110–116
25. Denu, J. M., and Dixon, J. E. (1995) A catalytic mechanism for the dual-specific phosphatases. *Proc. Natl. Acad. Sci. U. S. A.* **92**, 5910–5914
26. Zhang, Z. Y., Wang, Y., Wu, L., Fauman, E. B., Stuckey, J. A., Schubert, H. L., Saper, M. A., and Dixon, J. E. (1994) The Cys(X)5Arg catalytic motif in phosphoester hydrolysis. *Biochemistry* **33**, 15266–15270
27. Chen, P., Hutter, D., Yang, X., Gorospe, M., Davis, R. J., and Liu, Y. (2001) Discordance between the binding affinity of mitogen-activated protein kinase subfamily members for MAP kinase phosphatase-2 and their ability to activate the phosphatase catalytically. *J. Biol. Chem.* **276**, 29440–29449
28. Slack, D. N., Seternes, O. M., Gabrielsen, M., and Keyse, S. M. (2001) Distinct binding determinants for ERK2/p38 α and JNK map kinases mediate catalytic activation and substrate selectivity of map kinase phosphatase-1. *J. Biol. Chem.* **276**, 16491–16500
29. Nichols, A., Camps, M., Gillieron, C., Chabert, C., Brunet, A., Wilsbacher, J., Cobb, M., Pouyssegur, J., Shaw, J. P., and Arkininstall, S. (2000) Substrate recognition domains within extracellular signal-regulated kinase mediate binding and catalytic activation of mitogen-activated protein kinase phosphatase-3. *J. Biol. Chem.* **275**, 24613–24621
30. Favata, M. F., Horiuchi, K. Y., Manos, E. J., Daulerio, A. J., Stradley, D. A., Feeser, W. S., Van Dyk, D. E., Pitts, W. J., Earl, R. A., Hobbs, F., Copeland, R. A., Magolda, R. L., Scherle, P. A., and Trzaskos, J. M. (1998) Identification of a novel inhibitor of mitogen-activated protein kinase kinase. *J. Biol. Chem.* **273**, 18623–18632
31. Duncia, J. V., Santella, J. B., 3rd, Higley, C. A., Pitts, W. J., Wityak, J., Frieze, W. E., Rankin, F. W., Sun, J. H., Earl, R. A., Tabaka, A. C., Teleha, C. A., Blom, K. F., Favata, M. F., Manos, E. J., Daulerio, A. J., et al. (1998) MEK inhibitors: The chemistry and biological activity of U0126, its analogs, and cyclization products. *Bioorg. Med. Chem. Lett.* **8**, 2839–2844
32. Oho, M., Kinoshita, T., Okubo, M., Sato, K., Yamazaki, A., Arakawa, H., Nishimura, S., Inamura, N., Nakajima, H., Neya, M., Miyake, H., and Fujii, T. (2005) Identification of a selective ERK inhibitor and structural determination of the inhibitor-ERK2 complex. *Biochem. Biophys. Res. Commun.* **336**, 357–363
33. Haystead, T. A., Dent, P., Wu, J., Haystead, C. M., and Sturgill, T. W. (1992) Ordered phosphorylation of p42mapk by MAP kinase kinase. *FEBS Lett.* **306**, 17–22
34. Payne, D. M., Rossomando, A. J., Martino, P., Erickson, A. K., Her, J. H., Shabanowitz, J., Hunt, D. F., Weber, M. J., and Sturgill, T. W. (1991) Identification of the regulatory phosphorylation sites in pp42/mitogen-activated protein kinase (MAP kinase). *EMBO J.* **10**, 885–892
35. Torres, J. Z., Summers, M. K., Peterson, D., Brauer, M. J., Lee, J., Senese, S., Gholkar, A. A., Lo, Y. C., Lei, X., Jung, K., Anderson, D. C., Davis, D. P., Belmont, L., and Jackson, P. K. (2011) The STARD9/Kif16a kinesin associates with mitotic microtubules and regulates spindle pole assembly. *Cell* **147**, 1309–1323
36. Torres, J. Z., Miller, J. J., and Jackson, P. K. (2009) High-throughput generation of tagged stable cell lines for proteomic analysis. *Proteomics* **9**, 2888–2891
37. Bradley, M., Ramirez, I., Cheung, K., Gholkar, A. A., and Torres, J. Z. (2016) Inducible LAP-tagged stable cell lines for investigating protein function, spatiotemporal localization and protein interaction networks. *J. Vis. Exp.* <https://doi.org/10.3791/54870>
38. Gupta, G. D., Coyaud, E., Goncalves, J., Mojarad, B. A., Liu, Y., Wu, Q., Gheiratmand, L., Comartin, D., Tkach, J. M., Cheung, S. W., Bashkurov, M., Hasegan, M., Knight, J. D., Lin, Z. Y., Schueler, M., et al. (2015) A dynamic protein interaction landscape of the human centrosome-cilium interface. *Cell* **163**, 1484–1499
39. Kim, D. I., Jensen, S. C., Noble, K. A., Kc, B., Roux, K. H., Motamedchaboki, K., and Roux, K. J. (2016) An improved smaller biotin ligase for BioID proximity labeling. *Mol. Biol. Cell* **27**, 1188–1196
40. Patananan, A. N., Capri, J., Whitelegge, J. P., and Clarke, S. G. (2014) Non-repair pathways for minimizing protein isoaspartyl damage in the yeast *Saccharomyces cerevisiae*. *J. Biol. Chem.* **289**, 16936–16953
41. Stark, C., Breitkreutz, B. J., Reguly, T., Boucher, L., Breitkreutz, A., and Tyers, M. (2006) BioGRID: A general repository for interaction datasets. *Nucleic Acids Res.* **34**, D535–D539
42. Giurgiu, M., Reinhard, J., Brauner, B., Dunger-Kaltenbach, L., Fobo, G., Frishman, G., Montrone, C., and Ruepp, A. (2019) CORUM: The comprehensive resource of mammalian protein complexes-2019. *Nucleic Acids Res.* **47**, D559–D563
43. Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., Harris, M. A., Hill, D. P., Issel-Tarver, L., Kasarskis, A., Lewis, S., et al. (2000) Gene ontology: Tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* **25**, 25–29
44. Franz, M., Lopes, C. T., Huck, G., Dong, Y., Sumer, O., and Bader, G. D. (2016) Cytoscape.js: A graph theory library for visualisation and analysis. *Bioinformatics* **32**, 309–311
45. Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003) Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504
46. Cheung, K., Senese, S., Kuang, J., Bui, N., Ongpipattanakul, C., Gholkar, A., Cohn, W., Capri, J., Whitelegge, J. P., and Torres, J. Z. (2016) Proteomic analysis of the mammalian katanin family of microtubule-severing enzymes defines katanin p80 subunit B-like 1 (KATNBL1) as a regulator of mammalian katanin microtubule-severing. *Mol. Cell. Proteomics* **15**, 1658–1669
47. Gholkar, A. A., Senese, S., Lo, Y. C., Vides, E., Contreras, E., Hodara, E., Capri, J., Whitelegge, J. P., and Torres, J. Z. (2016) The X-linked-intellectual-disability-associated ubiquitin ligase Mid2 interacts with astrin and regulates astrin levels to promote cell division. *Cell Rep.* **14**, 180–188
48. Livak, K. J., and Schmittgen, T. D. (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2⁻(Delta Delta C(T)) method. *Methods* **25**, 402–408
49. Tighe, A., Johnson, V. L., and Taylor, S. S. (2004) Truncating APC mutations have dominant effects on proliferation, spindle checkpoint control, survival and chromosome stability. *J. Cell Sci.* **117**, 6339–6353
50. Rappsilber, J., Mann, M., and Ishihama, Y. (2007) Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nat. Protoc.* **2**, 1896–1906
51. Ishihama, Y., Oda, Y., Tabata, T., Sato, T., Nagasu, T., Rappsilber, J., and Mann, M. (2005) Exponentially modified protein abundance index (emPAI) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein. *Mol. Cell. Proteomics* **4**, 1265–1272

SHORT REPORT

The myosin regulatory light chain Myl5 localizes to mitotic spindle poles and is required for proper cell division

Ivan Ramirez¹ | Ankur A. Gholkar¹ | Erick F. Velasquez¹ | Xiao Guo¹ | Bobby Tofig² | Robert Damoiseaux^{2,3} | Jorge Z. Torres^{1,4,5} 

¹Department of Chemistry and Biochemistry, University of California, Los Angeles, California

²California NanoSystems Institute, Los Angeles, California

³Department of Molecular and Medical Pharmacology, Los Angeles, California

⁴Molecular Biology Institute, University of California, Los Angeles, California

⁵Jonsson Comprehensive Cancer Center, University of California, Los Angeles, California

Correspondence

Jorge Z. Torres, 607 Charles E. Young Drive East, Los Angeles, CA 90095, USA.

Email: torres@chem.ucla.edu

Funding information

National Institutes of Health, Grant/Award Number: P30CA016042; National Science Foundation, Grant/Award Number: MCB1912837

Abstract

Myosins are ATP-dependent actin-based molecular motors critical for diverse cellular processes like intracellular trafficking, cell motility, and cell invasion. During cell division, myosin MYO10 is important for proper mitotic spindle assembly, the anchoring of the spindle to the cortex, and positioning of the spindle to the cell mid-plane. However, myosins are regulated by myosin regulatory light chains (RLCs), and whether RLCs are important for cell division has remained unexplored. Here, we have determined that the previously uncharacterized myosin RLC Myl5 associates with the mitotic spindle and is required for cell division. We show that Myl5 localizes to the leading edge and filopodia during interphase and to mitotic spindle poles and spindle microtubules during early mitosis. Importantly, depletion of Myl5 led to defects in mitotic spindle assembly, chromosome congression, and chromosome segregation and to a slower transition through mitosis. Furthermore, Myl5 bound to MYO10 *in vitro* and co-localized with MYO10 at the spindle poles. These results suggest that Myl5 is important for cell division and that it may be performing its function through MYO10.

KEYWORDS

cell division, mitosis, mitotic spindle, Myl5, myosin light chain

1 | INTRODUCTION

The proper assembly of the bipolar mitotic microtubule spindle is critical to the fidelity of chromosome congression and segregation during cell division (Walczak & Heald, 2008). During development, the anchoring and positioning of the mitotic spindle regulates the establishment of the cell division plane that is critical for cell fate determination (Morin & Bellaiche, 2011). Important to mitotic spindle anchoring and positioning are astral microtubules that radiate out from the spindle poles and make contacts with the cell cortex (di Pietro, Echard, & Morin, 2016; Morin & Bellaiche, 2011; Nakajima, 2018). Abnormal spindle assembly and orientation can result in defective cell divisions that can lead to developmental and proliferative diseases (di Pietro et al., 2016; Morin & Bellaiche, 2011; Nakajima, 2018). Although numerous components involved in

assembling and orienting the spindle are known (Morin & Bellaiche, 2011; Walczak & Heald, 2008), the full complement of factors and the molecular signaling pathways that govern these events are not completely understood. Increasing evidence indicates that both the actin and microtubule cytoskeletal systems are necessary for proper cell division (Akhshi, Wernike, & Piekny, 2014; Kita et al., 2019). Although actin has been highly studied within the context of interphase cells where it establishes the cellular architecture and regulates numerous important processes like cell motility, intracellular trafficking, cell signaling pathways, and gene expression (Hyrskyluoto & Vartiainen, 2020; Moujaber & Stochaj, 2020; Svitkina, 2018; Titus, 2018), less is known about its role during early cell division. However, actin has been shown to be critical for anchoring the spindle through microtubule-actin interactions at the cell cortex, for spindle positioning at the mid plane, and for actomyosin

cellular constriction during cytokinesis (Akhshi et al., 2014; Chaigne et al., 2016; They & Bornens, 2006; Uraji, Scheffler, & Schuh, 2018). Additionally, evidence indicates that an actin mesh assembly supports the bipolar meiotic and mitotic spindles, where actin provides rigidity and aids in focusing the spindle (Kita et al., 2019; Mogessie & Schuh, 2017; Woolner, O'Brien, Wiese, & Bement, 2008; Wuhr, Mitchison, & Field, 2008). Therefore, actin plays an important role in ensuring the fidelity of cell division.

Myosins are ATP-dependent actin-based molecular motors that perform a variety of functions in muscle contraction, cargo transport, cell adhesion, and cell division; including spindle assembly, spindle orientation, and cytokinesis (Hartman & Spudich, 2012; Li, Lu, & Zhang, 2016). During cell division, myosin-II is critical for acto-myosin ring contraction during cytokinesis, which is essential for bisecting one cell into two daughter cells (Murthy & Wadsworth, 2005; Robinson & Spudich, 2000). Of interest, the unconventional myosin-10 (MYO10) has been shown to be an important factor for establishing the architecture and function of the mitotic spindle through its binding to both actin and microtubules (Kwon, Bagonis, Danuser, & Pellman, 2015; Sandquist, Larson, & Hine, 2016; Weber, Sokac, Berg, Cheney, & Bement, 2004; Woolner et al., 2008; Wuhr et al., 2008). MYO10 localizes to the spindle poles throughout mitosis and depletion of MYO10 leads to structural defects in the mitotic microtubule spindle, chromosome congression defects, and chromosome segregation defects (Kwon et al., 2008; Weber et al., 2004; Woolner et al., 2008; Wuhr et al., 2008). MYO10 also has important non-mitotic functions during interphase and in post-mitotic cells where it is important for filopodia formation and function (Kerber & Cheney, 2011; Quintero & Yengo, 2012; Sousa & Cheney, 2005). The filopodia core is composed of actin filaments (Leijnse, Oddershede, & Bendix, 2015) and MYO10 has been shown to be recruited to focal adhesions at the leading edge, undergoes intrafilopodial motility, accumulates at the filopodial tips, and promotes the formation and extension of filopodia (Berg & Cheney, 2002; Bohil, Robertson, & Cheney, 2006; He, Sakai, Tsukasaki, Watanabe, & Ikebe, 2017; Kerber et al., 2009; Sato et al., 2017; Tokuo, Mabuchi, & Ikebe, 2007). Therefore, myosins perform important functions that are necessary for cell division and cell motility.

The unconventional myosin holoenzymes typically consist of heavy and light chains (Li et al., 2016). Myosin light chains are required for the structural integrity of the myosin holoenzyme and have regulatory functions on the activity of the protein complex (Heissler & Sellers, 2014; Heissler & Sellers, 2016; Li et al., 2016). There are two major groups of myosin light chains, the essential light chains (ELCs) and the regulatory light chains (RLCs) (Heissler & Sellers, 2014; Heissler & Sellers, 2016). The ELCs are essential for the enzymatic activity of the myosin and removal or depletion of the ELCs from the myosin leads to a dramatic loss of myosin enzymatic activity (Heissler & Sellers, 2014; Heissler & Sellers, 2016). The RLCs are involved in regulating the enzymatic activity of the myosin and their removal or depletion typically leads to moderate effects on myosin activity (Heissler & Sellers, 2014; Heissler & Sellers, 2016). For example, both calmodulin and calmodulin-like protein have been shown to

be MYO10 light chains that regulate MYO10 motility and function (Bennett, Caride, Mauer, & Strehler, 2008; Bennett, Mauer, & Strehler, 2007; Homma, Saito, Ikebe, & Ikebe, 2001; Rogers & Strehler, 2001). Although myosin-II and MYO10 have important roles in cell division, the role of RLCs (if any) in cell division has remained unexplored.

We recently performed a microscopy-based RNAi screen using an siRNA library (25,620 siRNAs) targeting the druggable genome in HeLa cells, that included enzymes like kinases and components of molecular motors like myosins, to identify novel factors whose depletion led to a slowed cell division (Torres lab unpublished). This screen identified 13 novel proteins that were not known to be important for cell division, among which was myosin light chain 5 (Myl5). Although Myl5 has remained poorly characterized, based on its protein sequence similarity it is predicted to be a myosin RLC (Collins, Schappert, & Hayden, 1992). Dysregulation of *MYL5* mRNA levels has been observed in glioblastoma multiforme, cervical carcinoma, and breast cancer (Alshabi, Vastrad, Shaikh, & Vastrad, 2019; Savci-Heijink, Halfwerk, Koster, Horlings, & van de Vijver, 2019; Zhang et al., 2017). For example, *MYL5* mRNA levels are upregulated in late stage cervical cancer patients and is associated with poor survival (Zhang et al., 2017). Additionally, *MYL5* overexpression promoted tumor cell metastasis in a cervical cancer mouse model (Zhang et al., 2017). Here, we have discovered that Myl5 is important for mitotic spindle assembly, chromosome congression, and proper cell division. GFP-Myl5 localizes to the spindle poles during mitosis, indicating that its localization is cell cycle phase dependent. GFP-Myl5 co-localized with spindle pole proteins and MYO10 and bound to MYO10 in vitro. Importantly, depletion of Myl5 led to spindle assembly defects, chromosome congression defects, and chromosome segregation errors. These results suggest that Myl5 has important roles in spindle assembly and chromosome segregation and that it may be performing its function through its association with MYO10.

2 | RESULTS AND DISCUSSION

2.1 | In silico analysis of Myl5

Our recent genetic RNAi screen for novel cell division proteins led us to discover myosin light chain 5 (Myl5), an uncharacterized hypothetical myosin regulatory light chain (RLC) of the MLC2 type. Human Myl5 is a 173 amino acid protein with three EF hand domains predicted to be important for calcium binding in other myosin regulatory light chains (Figure 1a) (Grabarek, 2006; Heissler & Sellers, 2014). A phylogenetic tree (Hunt et al., 2018) analysis indicated that Myl5 was conserved among vertebrates (Figure 1b). An Online Mendelian Inheritance in Man search showed that the *MYL5* gene was within the 4p16.3 region where the Huntington Disease locus is located. However, *MYL5* has not been linked to inherited human diseases. Due to recent studies showing the upregulation of *MYL5* expression in cervical cancer (Zhang et al., 2017) and the down regulation of *MYL5* expression in breast (Savci-Heijink et al., 2019) and brain cancers

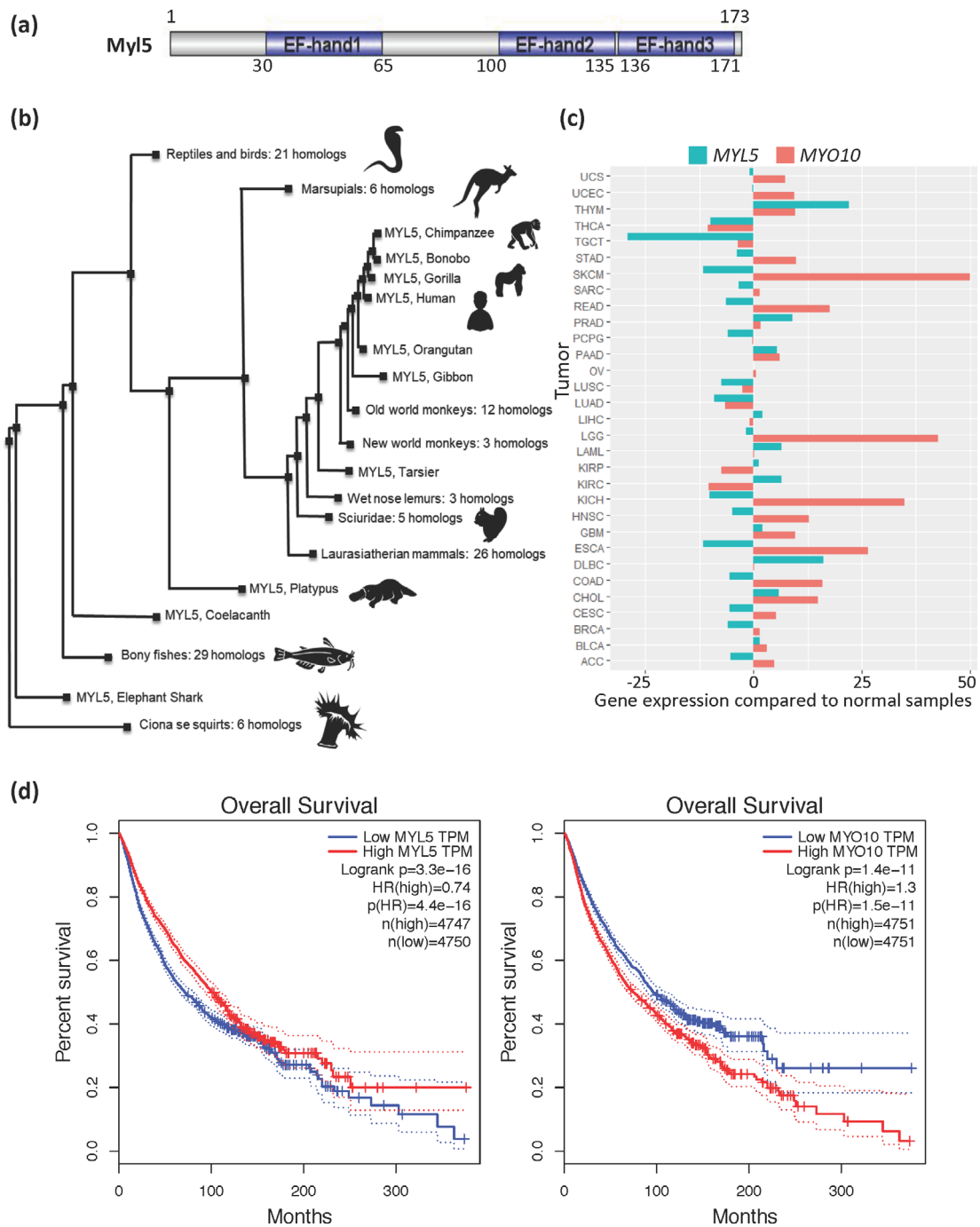


FIGURE 1 In silico analyses of Myl5. (a) Schematic of the human Myl5 (UniProtKB-Q02045) protein domain architecture with the EF hand domains highlighted in blue. The number of amino acid residues are indicated. (b) Phylogenetic tree analysis showing that Myl5 is conserved among vertebrates. (c) Analysis of MYL5 and MYO10 differential gene expression in a broad array of cancers versus matched normal samples with the Gene Expression Profiling and Interactive Analysis (GEPIA) web server. See Table S2 for a list of cancer types considered and their corresponding abbreviations. The median gene expression levels are on the x-axis and cancer type is on the y-axis. Positive values represent higher gene expression in tumor samples compared to normal samples and negative values represent lower gene expression in tumor samples compared to normal samples. (d) Correlation analysis between MYL5 and MYO10 mRNA levels and the overall survival of cancer patients using the GEPIA web server. Survival plots indicate time in months on the x-axis and percent survival on the y-axis. See materials and methods for analysis details [Color figure can be viewed at wileyonlinelibrary.com]

(Alshabi et al., 2019), we sought to determine if MYL5 was widely dysregulated in other types of cancers. Interestingly, analysis of MYL5 and MYO10 differential gene expression across a broad array of

cancers using the Gene Expression Profiling and Interactive Analysis (GEPIA) web server (Tang et al., 2017) showed that MYL5 mRNA levels were lower in most cancers compared to matched normal

samples, whereas *MYO10* mRNA levels were elevated in most cancers compared to normal samples (Figure 1c). Additionally, a GEPIA survival analysis showed that low *MYL5* mRNA levels and high *MYO10* mRNA levels related to an unfavorable overall survival (Figure 1d). Together, these analyses showed that the Myl5 protein is conserved among vertebrates and that *MYL5* and *MYO10* gene expression is widely dysregulated in cancer.

2.2 | Myl5 localizes to the spindle poles and spindle microtubules during cell division

Although previous genomic and bioinformatic studies had implicated Myl5 in myosin related functions and in tumorigenesis, its biological

function had remained poorly characterized. To begin to understand the cellular role of Myl5 and its link to tumorigenesis, we analyzed its subcellular localization throughout the cell cycle. First, we generated a LAP(GFP-TEV-S-Peptide)-Myl5 inducible stable cell line that expressed GFP-Myl5 upon induction with Dox (Figure S1a; Bradley, Ramirez, Cheung, Gholkar, & Torres, 2016; Torres, Miller, & Jackson, 2009). The LAP-Myl5 cell line was treated with Dox for 16 hr to express GFP-Myl5 and cells were fixed, stained with Hoechst 33342 DNA dye, and anti- α -Tubulin and anti-GFP antibodies, and imaged by immunofluorescence microscopy. During interphase GFP-Myl5 was dispersed throughout the nucleus and cytoplasm of the cell (Figure 2a). Interestingly, GFP-Myl5 localized to the spindle poles in early mitosis, and to a lesser extent the mitotic spindle, and remained associated with the poles until mitotic exit (Figure 2a). To further

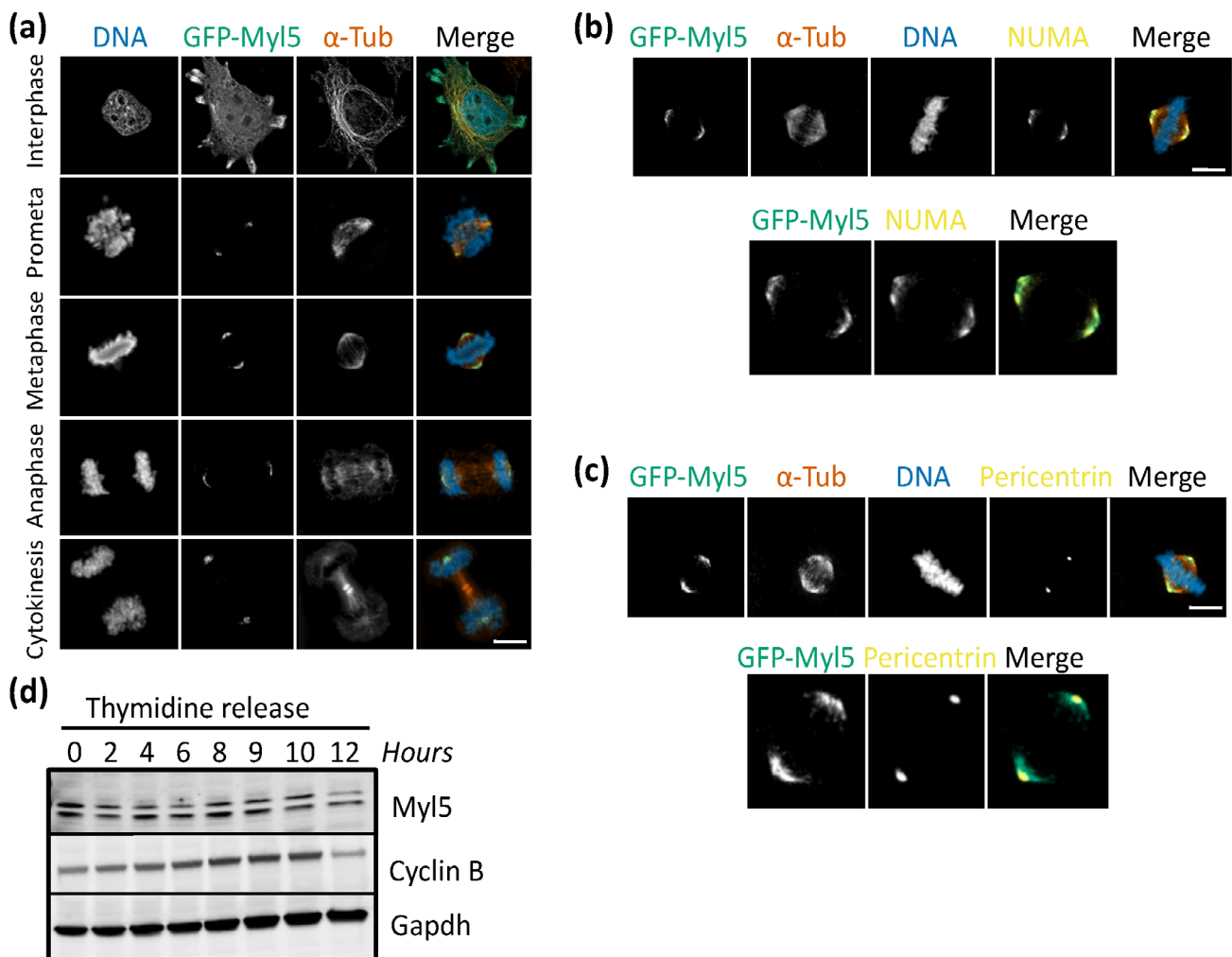


FIGURE 2 Myl5 localizes to mitotic spindle poles during mitosis. (a) The LAP (GFP-TEV-S-Peptide)-tagged-Myl5 HeLa inducible stable cell line was treated with Dox for 16 hr to express GFP-Myl5 and cells were fixed, stained with Hoechst 33342 DNA dye, and anti- α -Tubulin and anti-GFP antibodies, and imaged by immunofluorescence microscopy. Images show the cell cycle subcellular localization of GFP-Myl5 during interphase, prometaphase, metaphase, anaphase, and cytokinesis. Bar indicates 5 μ m. (b,c) Same as in (a), except that cells were also stained with anti-NUMA (b) or anti-Pericentrin (c) antibodies. Bar indicates 5 μ m. (d) Analysis of endogenous Myl5 protein levels throughout the cell cycle. HeLa cells were synchronized in G1/S, released into the cell cycle and cells were harvested at the indicated time points. Protein extracts were prepared, resolved by SDS-PAGE, transferred to a PVDF membrane and immunoblotted with the indicated antibodies. Gapdh is used a loading control [Color figure can be viewed at wileyonlinelibrary.com]

define the GFP-Myl5 subcellular localization in early mitosis, we performed immunofluorescence co-localization studies with centrosome and spindle pole markers. The GFP-Myl5 localization signal overlapped with NUMA at the spindle poles and encompassed the Pericentrin and Centrin signals, which stained the centrosomes (Figure 2b, c; Figure S1b). Furthermore, in cells with high levels of GFP-Myl5 expression, GFP-Myl5 also co-localized with TPX2 on the spindle microtubules (Figure S1c). Due to the change in GFP-Myl5 localization at mitotic entry, we next asked if endogenous Myl5 protein levels were also cell cycle regulated. HeLa cells were synchronized in G1/S with thymidine treatment, released into the cell cycle, cells were harvested every hour, and protein extracts were prepared. Immunoblot analysis of these samples with anti-Myl5 and anti-Cyclin B antibodies indicated that endogenous Myl5 protein levels remained steady in G1/S and G2/M and decreased slightly during mitotic exit, a time when mitotic Cyclin B levels decreased (Figure 2d). Additionally, the anti-Myl5 antibody recognized two protein bands that

corresponded with the size of the two Myl5 isoforms (19.5 kD, UniProtKB-Q02045-1; 14.9 kD, UniProtKB-Q02045-2). Together, these results indicated that the Myl5 protein is abundant throughout the cell cycle and that it undergoes a dynamic cell cycle dependent change in subcellular localization where it redistributes from the nucleus and cytoplasm in interphase to the spindle poles during mitotic entry and remains associated with the poles throughout mitosis.

2.3 | Myl5 is required for proper cell division

Next, we asked if Myl5 was required for cell division by depleting Myl5 in HeLa cells. First, we sought to identify siRNA oligonucleotides which reduced Myl5 protein levels to less than 10% compared to non-targeting control siRNA. HeLa cells were transfected with non-targeting control siRNA (siCtrl) or siRNAs targeting Myl5 (siM1-siM4) for 72 hr and cell

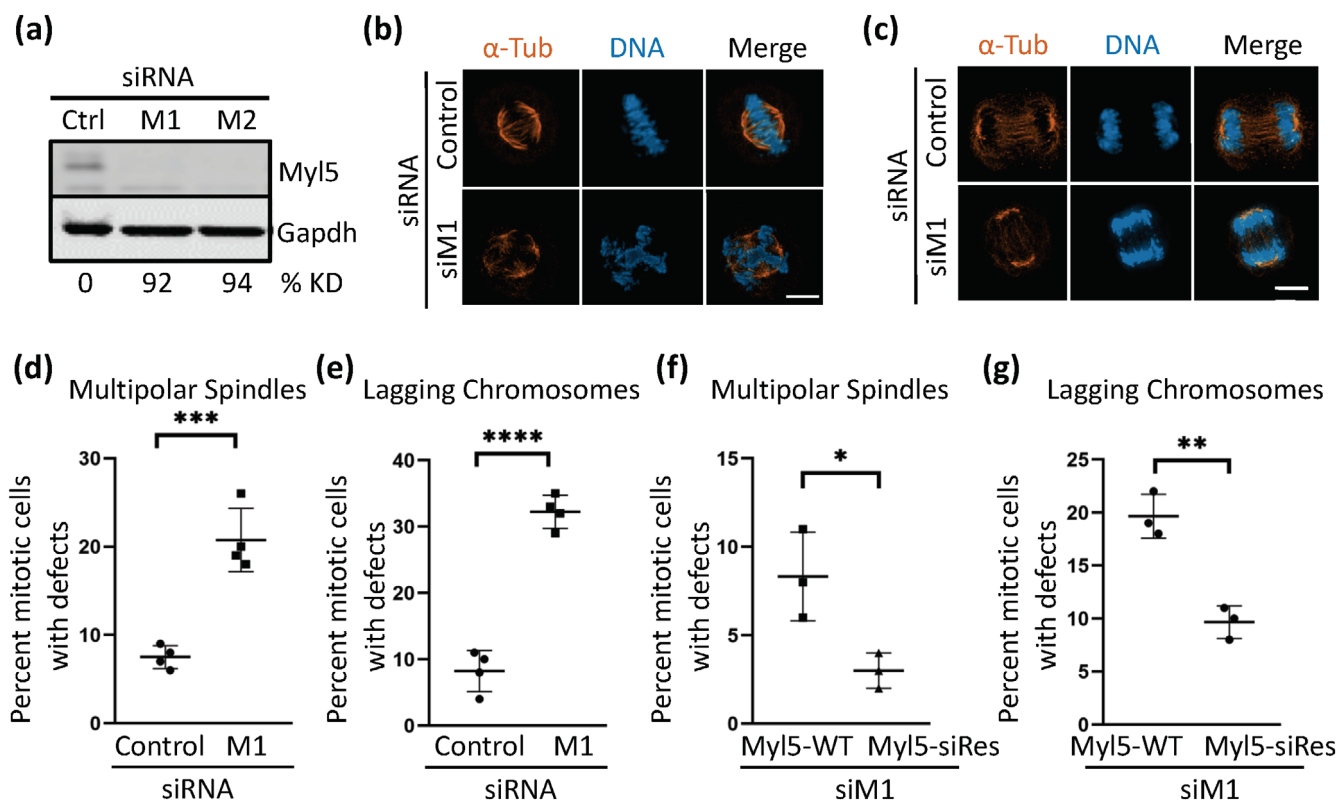


FIGURE 3 Depletion of Myl5 leads to spindle assembly and cell division defects. (a) Immunoblot analysis showing that siRNA oligonucleotides targeting MYL5 (M1 and M2) expression deplete Myl5 protein levels in HeLa cells compared to non-targeting control siRNA (siCtrl). Percent Myl5 protein level knockdown (% KD) normalized to Gapdh is indicated for each oligonucleotide. (b,c) Immunofluorescence microscopy of HeLa cells transfected with siCtrl or siM1 for 72 hr, fixed, and stained with Hoechst 33342 DNA dye and anti- α -Tubulin antibodies. Scale bar indicates 5 μ m. (d,e) Quantitation of the percent mitotic cells with multipolar spindles (d) and lagging chromosomes (e) in siCtrl or siM1 transfected cells. Data represent the average \pm SD of four independent experiments, 100 cells counted for each. *** indicates a p value = .0004 and **** a p value < .0001. (f-g) The LAP-Myl5-WT and LAP-Myl5-siRes (resistant to siM1 siRNA targeting Myl5) HeLa inducible stable cell lines were transfected with siM1 for 46 hr, synchronized in G1/S with thymidine for 18 hr, released into the cell cycle for 8 hr, and induced with Dox during the last 16 hr of the experiment to overexpress either GFP-Myl5-WT or GFP-Myl5-siRes. Cells were then fixed and analyzed by immunofluorescence microscopy and the percent mitotic cells with mitotic defects (multipolar spindles [f] and lagging chromosomes [g]) was quantified. Data represent the average \pm SD of three independent experiments, 100 cells counted for each. ** indicates a p value = .0026 and * indicates a p value = .0270 [Color figure can be viewed at wileyonlinelibrary.com]

lysates were prepared and analyzed by immunoblotting. The siM1-siM4 oligonucleotides depleted Myl5 protein levels to undetectable levels (Figure 3a; Figure S2a). Next, we sought to analyze the consequences of depleting Myl5 protein levels during cell division. HeLa cells were transfected with siCtrl or siM1-siM4 siRNAs for 72 hr. The cells were then fixed and co-stained with Hoechst 33342 (to visualize the DNA) and anti- α -Tubulin antibodies to detect the mitotic microtubule spindle. Interestingly, depletion of Myl5 led to a significant increase in cells with a defective mitosis, including an increase in the percentage of prometaphase cells with multipolar spindles (siM1 = 20.75 \pm 3.59%, p = .0004 compared to siCtrl = 7.5 \pm 1.29%) and anaphase cells with lagging chromosomes (siM1 = 32.25 \pm 2.5%, p < .0001 compared to siCtrl = 8.25 \pm 3.1%; Figure 3b-e; Figure S2b-f). Importantly, the mitotic defects (multipolar spindles and lagging chromosomes) observed upon siM1 treatment were rescued with the expression of a siRNA-resistant (siRes) version of GFP-Myl5 (for multipolar spindles—siM1+ GFP-Myl5-siRes = 3.0 \pm 1.0%, p = .0270 compared to siM1+ GFP-Myl5-WT = 8.33 \pm 2.51%; for lagging chromosomes—siM1+ GFP-Myl5-siRes = 9.67 \pm 1.5%, p = .0026 compared to siM1+ GFP-Myl5-WT = 19.67 \pm 2.08%) (Figure 3f,g; Figure S3). Together, these results indicated that Myl5 was required for proper cell division and that its depletion led to cell division errors.

2.4 | The levels of Myl5 affect the timing of cell division

Next, we asked if the overall time to cell division was affected by the depletion of Myl5. HeLa cells were transfected with non-targeting control siRNA (siCtrl) or siRNAs targeting Myl5 (siM1 and siM3) for

48 hr, synchronized in G1/S with thymidine treatment for 18 hr, and released in media containing the cell permeable DNA specific stain SiR-DNA (visible in the far-red channel). Five-hours post release live cells were imaged at $\times 20$ magnification at 5-min intervals for 18 hr. Movies were then analyzed and the time from nuclear DNA condensation to nuclear separation was quantified. This analysis showed that depletion of Myl5 led to a significant increase in the time that cells spent in mitosis with the average time from nuclear DNA condensation to nuclear separation for siM1 = 55 \pm 27.4 min (p < .0001) and siM3 = 46 \pm 22 min (p = .0288) compared to siCtrl = 38.9 \pm 21.6 min (Figure 4a,b; Videos S1-S3).

2.5 | Myl5 binds MYO10 and colocalizes with MYO10 to the spindle poles

Due to the ability of MYO10 to localize to the spindle poles in early mitosis and its functional importance in ensuring the fidelity of spindle assembly, chromosome congression, chromosome segregation, and cell division (Woolner et al., 2008), we asked if Myl5 and MYO10 shared a similar localization during mitosis. The LAP-Myl5 cell line was used to express GFP-Myl5 and cells were fixed, stained with Hoechst 33342 DNA dye, and anti- α -Tubulin, anti-MYO10, and anti-GFP antibodies, and imaged by immunofluorescence microscopy. Indeed, the GFP-Myl5 and MYO10 localization signals overlapped at the spindle poles throughout mitosis (Figure 5a). Next, we sought to determine if Myl5 and MYO10 could associate directly. In vitro binding experiments were performed with HA-Myl5 and FLAG-GFP or FLAG-MYO10. Indeed, MYO10 co-immunoprecipitated with Myl5 (Figure 5b). The IQ motifs of myosins are key sites for binding to

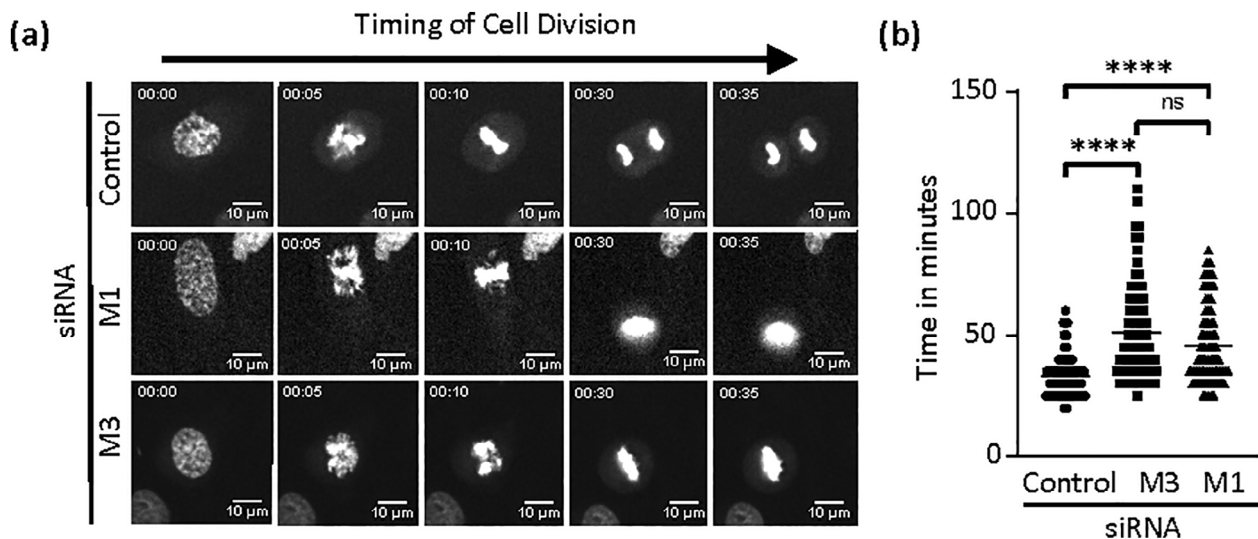


FIGURE 4 Modulation of Myl5 levels affects the time to cell division. (a) Live-cell time-lapse microscopy of HeLa cells transfected with siCtrl or siM1 or siM3 for 72 hr. Cells were then synchronized in G1/S with thymidine treatment for 18 hr, released, and imaged by staining the cells with SiR-DNA stain at 5 hr post-release. The indicated times are in minutes. See Videos S1-S3. (b) Quantitation of the time cells spend in mitosis from DNA condensation to chromosome separation. Y-axis indicates time in minutes. X-axis indicates the siRNA transfections. Data represent the average \pm SD of three independent experiments, 30 cells counted for each. **** indicates p value < .0001

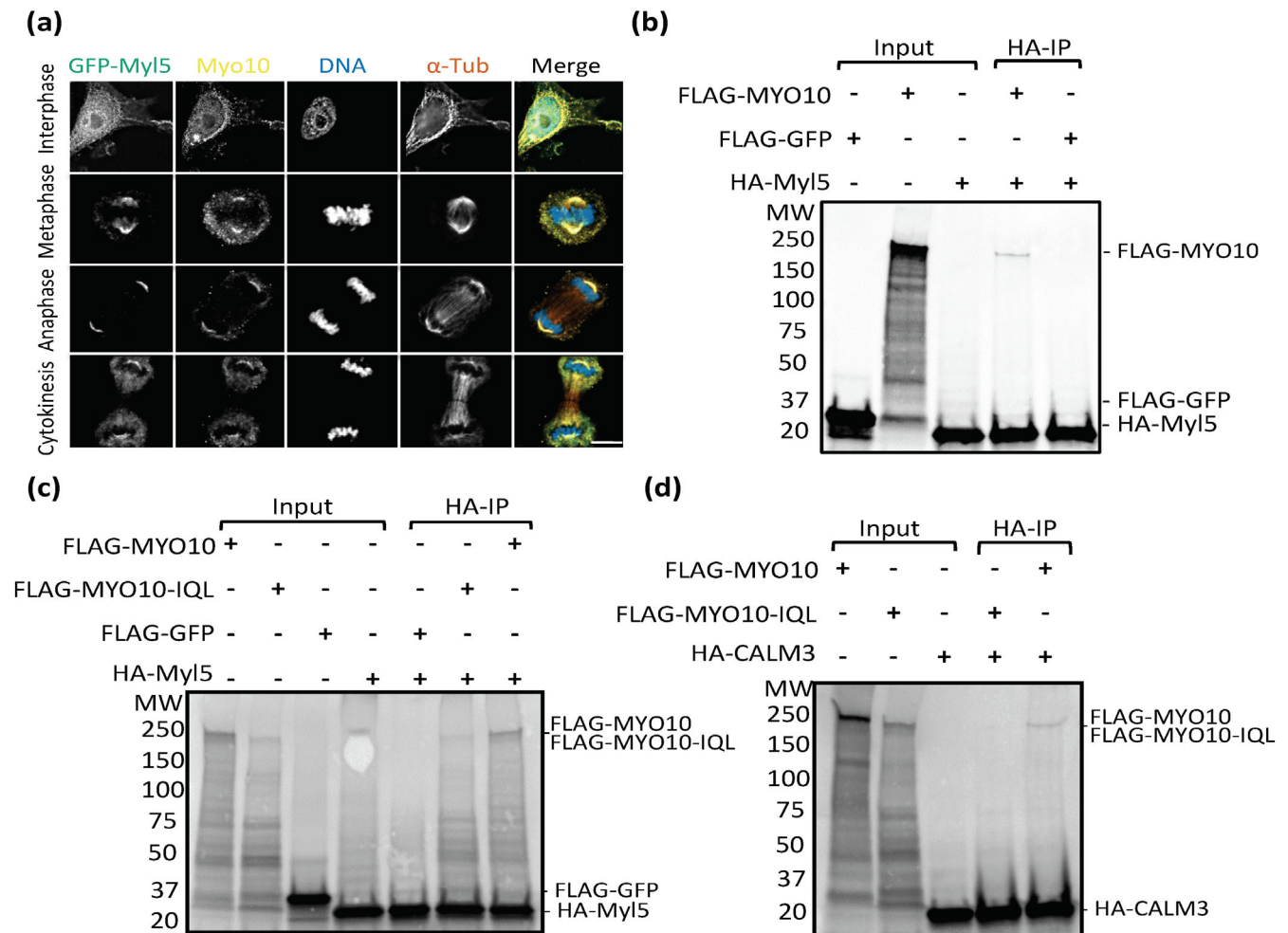


FIGURE 5 Myl5 colocalizes with MYO10 at mitotic spindle poles during mitosis and binds to MYO10 in vitro. (a) The LAP (GFP-TEV-S-Peptide)-tagged-Myl5 HeLa inducible stable cell line was treated with Dox for 16 hr to express GFP-Myl5 and cells were fixed, stained with Hoechst 33342 DNA dye, and anti- α -Tubulin and anti-MYO10 antibodies, and imaged by immunofluorescence microscopy. Images show the cell cycle subcellular localization of GFP-Myl5 and MYO10 during interphase, metaphase, anaphase, and cytokinesis. Bar indicates 5 μ m. (b) In vitro binding assays performed in the presence or absence of radiolabeled (35 S methionine) FLAG-MYO10, FLAG-GFP, or HA-Myl5. HA-Myl5 was immunoprecipitated (IP) and eluates were analyzed by radiometry. See Section 3 for experimental details. (c) Same as in (b), except that FLAG-MYO10-IQL (MYO10 IQ-less mutant) was added to the analysis. (d) Same as in (c), except that the binding of HA-CALM3 to FLAG-MYO10 or FLAG-MYO10-IQL was analyzed instead of HA-Myl5. See Figure S4 for control experiment showing that FLAG-MYO10 and FLAG-MYO10-IQL do not bind non-specifically to anti-HA beads [Color figure can be viewed at wileyonlinelibrary.com]

regulatory light chains (Heissler & Sellers, 2014), thus we asked if the IQ motifs of MYO10 were necessary for the MYO10-Myl5 interaction. To do this, we removed the three IQ motifs from MYO10 to generate MYO10 IQ-less (MYO10-IQL). In vitro binding experiments with HA-Myl5 or HA-CALM3 (Calmodulin 3, a known MYO10 regulatory light chain that binds to the MYO10 IQ motifs [Rogers & Strehler, 2001]) and FLAG-GFP, FLAG-MYO10, or FLAG-MYO10-IQL showed that Myl5 was able to bind to both MYO10 and MYO10-IQL, while CALM3 only bound to MYO10 and not MYO10-IQL (Figure 5c, d; Figure S4). Together, these data indicated that Myl5 binds to MYO10, independent of the IQ motifs, and colocalizes with MYO10 at the spindle poles during mitosis.

2.6 | Myl5 localizes to the leading edge of the cell and to filopodia in interphase cells

Within the context of cancer, actin-based structures like filopodia are critical for cell migration, invasion, and metastasis (Caswell & Zech, 2018; Jacquemet, Hamidi, & Ivaska, 2015). Of interest, MYO10 has a critical role in filopodia formation and accumulates at the tips of filopodia (Bohil et al., 2006; Kerber et al., 2009; Tokuo et al., 2007) and has been linked to promoting cancer invasion and metastasis, including in breast cancer and melanomas (Arjonen et al., 2014; Cao et al., 2014; Courson & Cheney, 2015; Tokuo, Bhawan, & Coluccio, 2018). Due to the association of Myl5 with MYO10, we

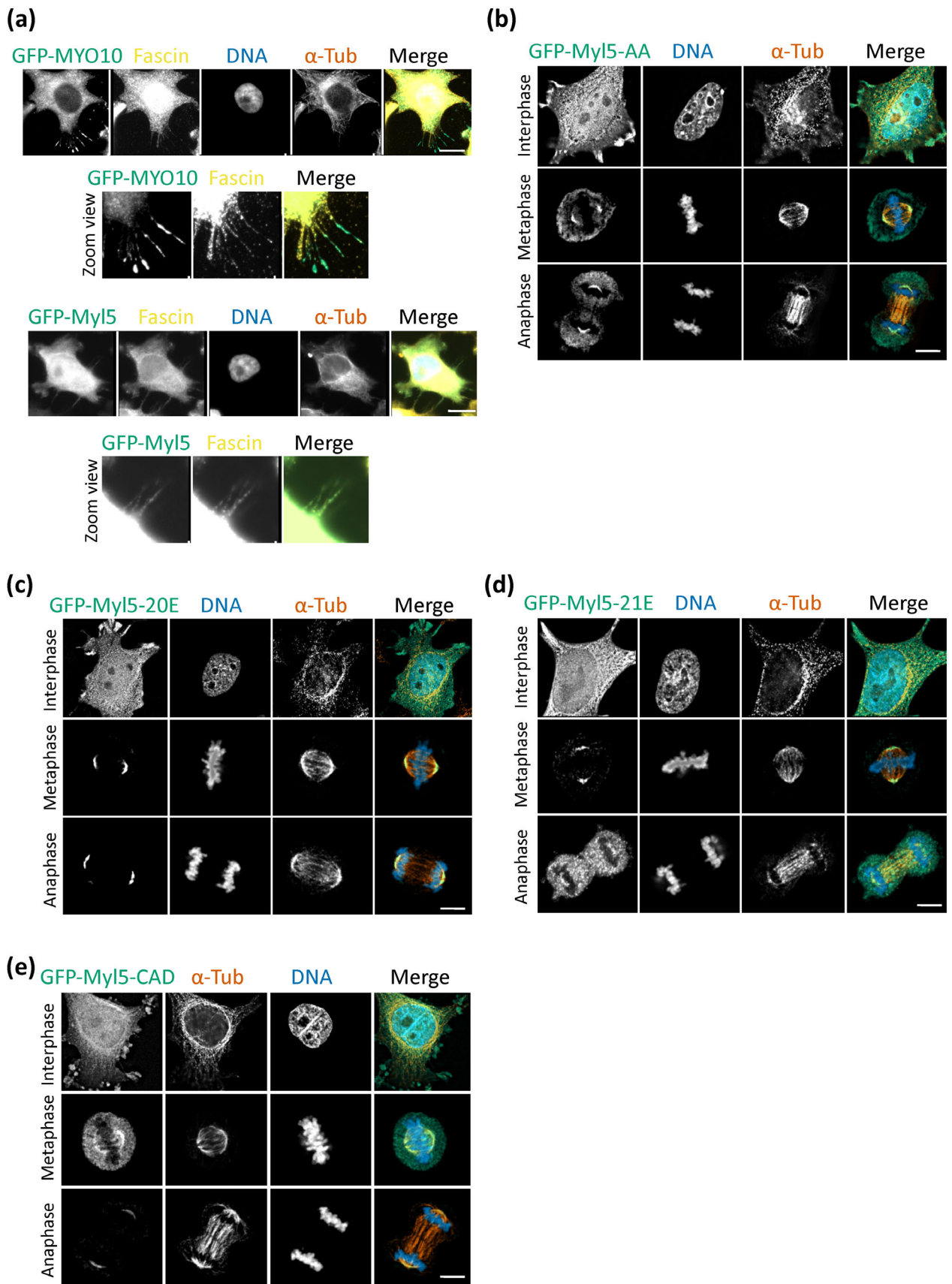


FIGURE 6 Legend on next page.

sought to determine whether Myl5 also localized to filopodia like MYO10. The LAP-Myl5 or LAP-MYO10 cell lines was used to express GFP-Myl5 or GFP-MYO10 and cells were fixed, stained with Hoechst 33342 DNA dye, anti- α -Tubulin, anti-Fascin (marker for filopodia [Edwards & Bryan, 1995; Otto, Kane, & Bryan, 1979]), and anti-GFP antibodies and imaged by immunofluorescence microscopy. Interestingly, GFP-Myl5 localized to the leading edge of the cell and throughout filopodia, but did not accumulate at the tips of filopodia like MYO10 (Figure 6a). Calcium binding and phosphorylation are two key mechanisms for regulating myosin RLC function (Heissler & Sellers, 2016). RLCs typically have multiple EF hands that are generally thought bind calcium, however, only EF hands with a calcium binding consensus pattern of amino acids are predicted to have the ability to bind calcium (prosite prorule annotation rule: PRU00448, [Sigrist et al., 2005]) (Grabarek, 2006). Specifically, Myl5 has three EF-hands, but only the N-terminal EF-hand contains the calcium binding consensus (amino acids 43–54; Figure 1a). Additionally, RLCs are typically phosphorylated at consensus Ser/Thr residues in their N-terminus, which are conserved in Myl5 (Figure S5; Yu, Chakravorty, Song, & Ferenczi, 2016). Therefore, we sought to determine whether deletion of the predicted Myl5 calcium binding site or mutation of the conserved RLC sites of phosphorylation (Serines 20 and 21, Figure S5) would perturb the localization of GFP-Myl5 to cytoskeletal structures. To do this, we generated LAP-tagged inducible stable cell lines of Myl5 calcium binding site deletion (CAD), phospho-null (20A and 21A [AA]) and phospho-mimic (20E or 21E) mutants; although phospho-mimetic mutants in other RLCs do not always have the anticipated effect on myosin activity (Heissler & Sellers, 2015; Vasquez, Heissler, Billington, Sellers, & Martin, 2016). Immunofluorescence microscopy of HeLa cells overexpressing these GFP-Myl5 mutants, showed that they were capable of localizing to the leading edge, filopodia, and the spindle poles, similar to wild type GFP-Myl5 (Figure 6b–e). Additionally, no major perturbations to the overall architecture of filopodia, spindles, and spindle poles were observed by the overexpression of these Myl5 mutants. Together, these results showed that GFP-Myl5 localizes to the leading edge and filopodia during interphase and that conserved residues in Myl5 that regulate the function of other RLCs are not required for its localization to these structures.

2.7 | Conclusion

Although actin and the unconventional myosin MYO10 had been implicated in ensuring the fidelity of mitotic spindle assembly and cell division, the role of myosin RLCs during cell division remained

unknown. Here, we have determined that Myl5 is a novel and important factor necessary for proper cell division. GFP-Myl5 localized to the leading edge and filopodia in interphase cells and to the spindle poles and spindle microtubules during early mitosis. Depletion of Myl5 led to mitotic spindle defects, errors in chromosome congression and segregation, and a slowed progression through mitosis. These Myl5 depletion phenotypes were similar to those reported upon MYO10 depletion (Woolner et al., 2008), albeit less severe. Interestingly, the GFP-Myl5 immunofluorescence signal overlapped with MYO10 at the spindle poles throughout mitosis and Myl5 bound directly to MYO10 *in vitro*. Our results suggest that Myl5 is important for cell division and that it may function through MYO10. To our knowledge, Myl5 is the first myosin RLC family member that has been implicated in mitotic spindle assembly.

Of the ~40 myosins encoded in the human genome, at least 10 (including MYO10) have been implicated in tumorigenesis (Li & Yang, 2016). Our analysis showing that *MYL5* and *MYO10* gene expression are frequently dysregulated in cancer compared to normal samples is intriguing and puzzling. While *MYL5* mRNA levels were lower in most cancers and related to unfavorable survival, *MYO10* mRNA levels were elevated in most cancers and related to unfavorable survival (Figure 1c,d). These results are consistent with reports showing low levels of *MYL5* mRNA in breast (Savci-Heijink et al., 2019) and brain cancers (Alshabi et al., 2019) and high *MYO10* mRNA levels in invasive and metastatic breast cancer and melanoma (Arjonen et al., 2014; Cao et al., 2014; Courson & Cheney, 2015; Tokuo et al., 2018). However, others have reported that *MYL5* mRNA levels are elevated in late stage cervical cancer patients, are associated with poor survival, and can promote tumor cell metastasis in mouse models of cervical cancer (Zhang et al., 2017). Similarly, the upregulation of *MYO10* mRNA levels in invasive and metastatic cancers is consistent with its critical role in promoting filopodia formation, which are important for cell motility and invasion (Berg & Cheney, 2002; Bohil et al., 2006; Caswell & Zech, 2018; He et al., 2017; Jacquemet et al., 2015; Kerber et al., 2009; Sato et al., 2017; Tokuo et al., 2007). However, it is the depletion of MYO10 that leads to cell division defects and genetic instability (Kwon et al., 2008; Weber et al., 2004; Woolner et al., 2008; Wuhr et al., 2008) and what role this may play in early stage cancers remains to be determined. Therefore, it is possible that *MYL5* and *MYO10* mRNA levels are differentially expressed in early versus late stage tumors and further research in this area is warranted. Although our data suggest that Myl5 may be affecting cell division through MYO10, it is possible that Myl5 may function independently or with other non-motor proteins. We also note that during interphase, myosin light

FIGURE 6 Myl5 localizes to the leading edge and filopodia. (a) The LAP (GFP-TEV-S-Peptide)-tagged-Myl5 or MYO10 HeLa inducible stable cell lines were treated with Dox for 16 hr to express GFP-Myl5 or GFP-MYO10 and cells were fixed, stained with Hoechst 33342 DNA dye, anti-GFP and anti- α -Tubulin, and anti-Fascin (filopodia marker) antibodies, and imaged by immunofluorescence (IF) microscopy. Bars indicate 5 μ m. Bottom panels of GFP-Myl5 and GFP-MYO10 IF images show magnified view of filopodia. (b–e) GFP-Myl5-AA (phospho-null), GFP-Myl5-20E and GFP-Myl5-21E (phospho-mimics), and GFP-Myl5-CAD (calcium-binding domain deletion) mutants were expressed in HeLa cells for 16 hr and cells were fixed, stained with Hoechst 33342 DNA dye, and anti- α -Tubulin and anti-GFP antibodies and imaged by IF microscopy. Bars indicate 5 μ m [Color figure can be viewed at wileyonlinelibrary.com]

chains have been implicated in regulating gene expression by binding to specific sequences within the promoter region of target genes (Li & Sarna, 2009; Zhang et al., 2015) and that Myl5 has been shown to bind the promoter region of HIF-1 α , an important factor in tumorigenesis, and regulates its expression (Rankin & Giaccia, 2016; Zhang et al., 2017). Consistent with this function, GFP-Myl5 localized to both the cytoplasm and nucleus in interphase cells (Figure 6a). Therefore in addition to its cytoskeleton-related function, Myl5 has cytoskeleton unrelated functions in gene expression that may contribute to tumorigenesis.

3 | MATERIALS AND METHODS

3.1 | Cell culture

HeLa cells were grown in F12:DMEM 50:50 (Hyclone, Logan UT, USA) with 10% FBS, 2 mM L-glutamine and antibiotics in 5% CO₂ at 37°C. Cells were synchronized in G1/S by treatment with 2 mM thymidine (Sigma-Aldrich, St. Louis MO, USA) for 18 hr. The following siRNAs were used for siRNA transfections: ThermoFisher (Waltham MA, USA) Silencer Select 4390843 (control non-targeting siRNA) and S9187 and S9188 (M1 and M2 siRNAs targeting MYL5); Dharmacon (Lafayette CO, USA) ON-TARGETplus D-001810-10 (control non-targeting siRNA) and J-011739-03 and J-011739-04 (M3 and M4 siRNAs targeting MYL5) were used as described previously (Torres, Ban, & Jackson, 2010). See Table S1 for a list of key reagents and resources used in this study and their pertinent information.

3.2 | Generation of the LAP-Myl5 inducible stable cell line

The HeLa LAP(GFP-TEV-S-Peptide)-Myl5, -Myl5-AA, -Myl5-20E, -Myl5-21E, -Myl5-CAD, -Myl5-siRes, and -MYO10 inducible stable cell lines were generated as described previously (Bradley et al., 2016; Torres et al., 2009). Briefly, full-length MYL5 (coding for amino acid residues 1–173) and mutant derivatives (alanine mutations at both Ser20 and Ser21 [AA]; Glu mutations at either Ser20 [20E] or Ser21 [21E]; deletion of the calcium binding domain [amino acids 43–54, CAD]) and full-length MYO10 (coding for amino acid residues 1–2058) were cloned into pDONR221 and transferred to pGLAP1 through a Gateway reaction to generate the pGLAP1 vectors with these ORFs that were transfected into HeLa Flp-In T-Rex cells to generate their respective inducible stable cell lines.

3.3 | Immunoblotting

For Myl5 cell cycle protein expression analysis, HeLa cells were synchronized in G1/S with 2 mM thymidine for 18 hr. Cells were then washed with PBS three times and twice with F12:DMEM media with 10% FBS and released into the cell cycle. Cells were harvested at the

indicated time points, lysed, and protein extracts were resolved on a 4–20% SDS-PAGE and transferred to a PVDF membrane. The membranes were immunoblotted with the indicated antibodies and imaged with a LiCOR (Lincoln NE, USA) system. The same approach was used to detect Myl5 protein depletion upon siRNA transfections without the cell synchronization step. Cell extract preparation and immunoblot analyses with the indicated antibodies were as described previously (Gholkar et al., 2016).

3.4 | Fixed-cell immunofluorescence microscopy and live-cell time-lapse microscopy

Fixed-cell immunofluorescence microscopy was performed as described previously (Gholkar et al., 2016). Briefly, non-transfected cells or cells that had been transfected with the indicated siRNAs for 48 hr were arrested in G1/S with 2 mM thymidine for 18 hr, washed, and released into fresh media for 8 hr. Cells were then fixed with 4% paraformaldehyde, permeabilized with 0.2% Triton X-100/PBS, and co-stained with 0.5 μ g/ml Hoechst 33342 (ThermoFisher) to visualize the DNA and the indicated antibodies. A Leica (Wetzlar, Germany) DMI6000 microscope (Leica DFC360 FX Camera, 63x/1.40–0.60 NA oil objective, Leica AF6000 software) was then used to capture the images, which were deconvolved with the Leica Application Suite 3D Deconvolution software and exported as TIFF files. For quantifying mitotic defects, the data from four independent experiments, with 100 cells counted for each, was used to quantify the average \pm SD. For time-lapse microscopy, HeLa cells were transfected with the indicated siRNAs for 48 hr, arrested in G1/S with 2 mM thymidine for 18 hr, washed, and released into fresh media containing 100 nM SiR-DNA stain (Cytoskeleton Inc., Denver CO, USA). Cells were imaged live 5-hr post release for 18 hr using an ImageXpress XL imaging system (Molecular Devices, San Jose CA, USA) with a \times 20 air objective at 37°C in 5% CO₂. Captured images were exported as a video at one-half frames per second using Image J and the videos were saved as AVI movies. Each frame represents a 5-min interval. For quantifying the timing of cell division, the data from three independent experiments, with 30 cells counted for each, was used to quantify the average time in minutes from DNA condensation for nuclear separation \pm SD.

3.4.1 | Statistical analysis

All statistical data are presented as the average \pm SD from at least three independent experiments. Outliers were considered in time-lapse experiments by using Tukey's method in R (<https://www.r-project.org/>). For experiments where two groups were compared, they were analyzed using unpaired Student's *t* test. Data was judged to be statistically significant when $p < .05$. For experiments where three or more groups were compared, they were first tested for significance using ANOVA statistical test. If *p*-value showed significance ($p < .05$), multiple pair-wise comparisons were performed between

the means of groups using Tukey Honest Significant Difference and Dunnett's tests. All statistical figures were generated with GraphPad (La Jolla CA, USA) Prism 5.

3.5 | Generation of plasmids and in vitro binding assays

For in vitro binding assays, full-length human *MYL5* (encoding amino acid residues 1–173) or *CALM3* (encoding amino acid residues 1–149) were fused to the C-terminus of the HA-tag to generate the pCS2-HA-*MYL5* and pCS2-HA-*CALM3* vectors. Similarly full-length *MYO10* (encoding amino acid residues 1–2058) or *MYO10* lacking amino acids 742–817 that contain the IQ motifs (*MYO10-IQL*) were fused to the C-terminus of the FLAG-tag to generate the pCS2-FLAG-*MYO10* and pCS2-FLAG-*MYO10-IQL* vectors. In vitro binding assays were performed as described previously (Gholkar et al., 2016). Briefly, HA-My15, HA-CALM3, FLAG-MYO10, FLAG-MYO10-IQL, and FLAG-GFP (negative control) were in vitro transcribed and translated (IVT) using TNT[®] Quick Coupled Transcription/Translation System, (Promega, Madison WI, USA) in 10 μ l reactions. Magnetic HA beads (MBL International, Woburn MA, USA) were washed three times and equilibrated with wash buffer (50 mM Tris pH 7.4, 200 mM KCl, 1 mM DTT, 0.5% NP-40, and Halt Protease and Phosphatase Inhibitor Cocktail). IVT reactions were added to the equilibrated HA beads and incubated for 1.5 hr at 30°C with gentle shaking. Beads were washed three times with wash buffer and eluted by boiling for 5 min with 2X Laemmli SDS sample buffer. Samples were resolved on a 4–20% gradient Tris gel with Tris-Glycine SDS running buffer, transferred to an Immobilon PVDF membrane (EMD Millipore, Burlington MA, USA), and membranes were analyzed with a PharosFX Plus molecular imaging system (Bio-Rad, Hercules CA, USA).

3.6 | In silico analysis of My15

The My15 phylogenetic tree was constructed by querying Ensembl (Hunt et al., 2018) (<https://www.ensembl.org/>) for My15 (ID: ENSGT00940000163023) and Figure 1b was generated by reconstructing the phylogenetic tree on Ensembl using images from LogoMarkr (<https://logomakr.com/>). For analysis of *MYL5* and *MYO10* differential gene expression in cancer cells compared to normal counterparts, their gene expression profiles were retrieved from the GEPIA, an interactive web server for cancer genomics that compares cancer and normal gene expression (Tang et al., 2017). For a list of cancer types considered and their corresponding abbreviations refer to Table S2. The median gene expression for both tumor and matched normal samples were compared by subtracting median normal from tumor gene expression for each cancer type. Positive values represent higher gene expression in tumor samples compared to normal samples and negative values represent lower gene expression in tumor samples compared to normal samples. Figure 1c summarizes the gene expression compared to normal samples in different tumor types for

both *MYL5* and *MYO10*. Correlation analysis between *MYL5* and *MYO10* mRNA levels and the overall survival of cancer patients was carried out using the Survival Plot tool in GEPIA with default parameters. Datasets for all cancers and matched normal samples listed on Table S2 were used in the analysis. Survival plots were exported as PDF files.

3.7 | Antibodies

Immunofluorescence and immunoblotting were carried out using antibodies against: My15, MYO10, and Fascin (Proteintech (Rosemont IL, USA): 14249-1-AP, 24565-1-AP, and 66321-1-1g); Pericentrin (Novus Biologicals (Centennial CO, USA): NB-100-68277); GFP (Abcam (Cambridge, United Kingdom): ab13970); Gapdh (GeneTex (Irvine CA, USA) GTX100118); α -Tubulin (Bio-Rad (Hercules CA, USA): MCA78G); Cyclin B (Santa Cruz (Dallas TX, USA): sc-245). Centrin antibodies were a gift from J. Salisbury and NUMA and TPX2 antibodies were gifts from D. Compton. Secondary antibodies conjugated to FITC, Cy3, and Cy5 were from Jackson Immuno Research (West Grove PA, USA) and those conjugated to IRDye 680 and IRDye 800 were from LI-COR Biosciences (Lincoln NE, USA).

ACKNOWLEDGMENTS

This material is based upon work supported by the National Science Foundation under Grant Number MCB1912837 to J.Z.T., any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. This work was supported in part by a grant to The University of California, Los Angeles from the Howard Hughes Medical Institute through the James H. Gilliam Fellowships for Advanced Study program to E.F.V. This work was also supported by an NSF Louis Stokes Alliances for Minority Participation Bridge to the Doctorate Fellowship and Cota Robles Fellowship to I.R. Work performed in the UCLA Molecular Screening Shared Resource was supported by the National Cancer Institute of the National Institutes of Health under award number P30CA016042.

AUTHOR CONTRIBUTIONS

Ivan Ramirez, Xiao Guo, Erick F. Velasquez, Ankur A. Gholkar, Bobby Tofig, Robert Damoiseaux, and Jorge Z. Torres contributed to the design, execution, and analysis of experiments and data. Ivan Ramirez and Jorge Z. Torres wrote the manuscript with input and editing from all authors. All authors read and approved the final version of the manuscript.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon request.

ORCID

Jorge Z. Torres  <https://orcid.org/0000-0002-2158-889X>

REFERENCES

- Akhshi, T. K., Wernike, D., & Piekny, A. (2014). Microtubules and actin crosstalk in cell migration and division. *Cytoskeleton (Hoboken)*, 71(1), 1–23.
- Alshabi, A. M., Vastrad, B., Shaikh, I. A., Vastrad, C. (2019). Identification of crucial candidate genes and pathways in glioblastoma multiform by bioinformatics analysis. *Biomolecules*, 9(5), 201. <http://doi.org/10.3390/biom9050201>.
- Arjonen, A., Kaukonen, R., Mattila, E., Rouhi, P., Hognas, G., Sihto, H., et al. (2014). Mutant p53-associated myosin-X upregulation promotes breast cancer invasion and metastasis. *The Journal of Clinical Investigation*, 124(3), 1069–1082.
- Bennett, R. D., Caride, A. J., Mauer, A. S., & Strehler, E. E. (2008). Interaction with the IQ3 motif of myosin-10 is required for calmodulin-like protein-dependent filopodial extension. *FEBS Letters*, 582(16), 2377–2381.
- Bennett, R. D., Mauer, A. S., & Strehler, E. E. (2007). Calmodulin-like protein increases filopodia-dependent cell motility via up-regulation of myosin-10. *The Journal of Biological Chemistry*, 282(5), 3205–3212.
- Berg, J. S., & Cheney, R. E. (2002). Myosin-X is an unconventional myosin that undergoes intrafilopodial motility. *Nature Cell Biology*, 4(3), 246–250.
- Bohil, A. B., Robertson, B. W., & Cheney, R. E. (2006). Myosin-X is a molecular motor that functions in filopodia formation. *Proceedings of the National Academy of Sciences of the United States of America*, 103(33), 12411–12416.
- Bradley, M., Ramirez, I., Cheung, K., Gholkar, A. A., & Torres, J. Z. (2016). Inducible LAP-tagged stable cell lines for investigating protein function, spatiotemporal localization and protein interaction networks. *Journal of Visualized Experiments*, 118(118), 54870.
- Cao, R., Chen, J., Zhang, X., Zhai, Y., Qing, X., Xing, W., ... Zhu, X. (2014). Elevated expression of myosin X in tumours contributes to breast cancer aggressiveness and metastasis. *British Journal of Cancer*, 111(3), 539–550.
- Caswell, P. T., & Zech, T. (2018). Actin-based cell protrusion in a 3D matrix. *Trends in Cell Biology*, 28(10), 823–834.
- Chaigne, A., Campillo, C., Voituriez, R., Gov, N. S., Sykes, C., Verlhac, M. H., & Terret, M. E. (2016). F-actin mechanics control spindle centring in the mouse zygote. *Nature Communications*, 7, 10253.
- Collins, C., Schappert, K., & Hayden, M. R. (1992). The genomic organization of a novel regulatory myosin light chain gene (MYL5) that maps to chromosome 4p16.3 and shows different patterns of expression between primates. *Human Molecular Genetics*, 1(9), 727–733.
- Courson, D. S., & Cheney, R. E. (2015). Myosin-X and disease. *Experimental Cell Research*, 334(1), 10–15.
- di Pietro, F., Echard, A., & Morin, X. (2016). Regulation of mitotic spindle orientation: An integrated view. *EMBO Reports*, 17(8), 1106–1130.
- Edwards, R. A., & Bryan, J. (1995). Fascins, a family of actin bundling proteins. *Cell Motility and the Cytoskeleton*, 32(1), 1–9.
- Gholkar, A. A., Senese, S., Lo, Y. C., Vides, E., Contreras, E., Hodara, E., ... Torres, J. Z. (2016). The X-linked-intellectual-disability-associated ubiquitin ligase Mid2 interacts with astrin and regulates astrin levels to promote cell division. *Cell Reports*, 14(2), 180–188.
- Grabarek, Z. (2006). Structural basis for diversity of the EF-hand calcium-binding proteins. *Journal of Molecular Biology*, 359(3), 509–525.
- Hartman, M. A., & Spudich, J. A. (2012). The myosin superfamily at a glance. *Journal of Cell Science*, 125(Pt 7), 1627–1632.
- He, K., Sakai, T., Tsukasaki, Y., Watanabe, T. M., & Ikebe, M. (2017). Myosin X is recruited to nascent focal adhesions at the leading edge and induces multi-cycle filopodial elongation. *Scientific Reports*, 7(1), 13685.
- Heissler, S. M., & Sellers, J. R. (2014). Myosin light chains: Teaching old dogs new tricks. *BioArchitecture*, 4(6), 169–188.
- Heissler, S. M., & Sellers, J. R. (2015). Four things to know about myosin light chains as reporters for non-muscle myosin-2 dynamics in live cells. *Cytoskeleton (Hoboken)*, 72(2), 65–70.
- Heissler, S. M., & Sellers, J. R. (2016). Various themes of myosin regulation. *Journal of Molecular Biology*, 428(9 Pt B), 1927–1946.
- Homma, K., Saito, J., Ikebe, R., & Ikebe, M. (2001). Motor function and regulation of myosin X. *The Journal of Biological Chemistry*, 276(36), 34348–34354.
- Hunt, S. E., McLaren, W., Gil, L., Thormann, A., Schuilenburg, H., Sheppard, D., ... Cunningham, F. (2018). Ensembl variation resources. *Database*, 2018. <http://dx.doi.org/10.1093/database/bay119>.
- Hyrskyluoto, A., & Vartiainen, M. K. (2020). Regulation of nuclear Actin dynamics in development and disease. *Current Opinion in Cell Biology*, 64, 18–24.
- Jacquemet, G., Hamidi, H., & Ivaska, J. (2015). Filopodia in cell adhesion, 3D migration and cancer cell invasion. *Current Opinion in Cell Biology*, 36, 23–31.
- Kerber, M. L., & Cheney, R. E. (2011). Myosin-X: A MyTH-FERM myosin at the tips of filopodia. *Journal of Cell Science*, 124(Pt 22), 3733–3741.
- Kerber, M. L., Jacobs, D. T., Campagnola, L., Dunn, B. D., Yin, T., Sousa, A. D., ... Cheney, R. E. (2009). A novel form of motility in filopodia revealed by imaging myosin-X at the single-molecule level. *Current Biology*, 19(11), 967–973.
- Kita, A. M., Swider, Z. T., Erofeev, I., Halloran, M. C., Goryachev, A. B., & Bement, W. M. (2019). Spindle-F-actin interactions in mitotic spindles in an intact vertebrate epithelium. *Molecular Biology of the Cell*, 30(14), 1645–1654.
- Kwon, M., Bagonis, M., Danuser, G., & Pellman, D. (2015). Direct microtubule-binding by myosin-10 orients centrosomes toward retraction fibers and subcortical Actin clouds. *Developmental Cell*, 34(3), 323–337.
- Kwon, M., Godinho, S. A., Chandhok, N. S., Ganem, N. J., Azioune, A., Thery, M., & Pellman, D. (2008). Mechanisms to suppress multipolar divisions in cancer cells with extra centrosomes. *Genes & Development*, 22(16), 2189–2203.
- Leijnse, N., Oddershede, L. B., & Bendix, P. M. (2015). An updated look at actin dynamics in filopodia. *Cytoskeleton (Hoboken)*, 72(2), 71–79.
- Li, J., Lu, Q., & Zhang, M. (2016). Structural basis of cargo recognition by unconventional myosins in cellular trafficking. *Traffic*, 17(8), 822–838.
- Li, Q., & Sarna, S. K. (2009). Nuclear myosin II regulates the assembly of preinitiation complex for ICAM-1 gene transcription. *Gastroenterology*, 137(3), 1051–1060 e1-3.
- Li, Y. R., & Yang, W. X. (2016). Myosins as fundamental components during tumorigenesis: Diverse and indispensable. *Oncotarget*, 7(29), 46785–46812.
- Mogessie, B., & Schuh, M. (2017). Actin protects mammalian eggs against chromosome segregation errors. *Science*, 357(6353), eaal1647.
- Morin, X., & Bellaiche, Y. (2011). Mitotic spindle orientation in asymmetric and symmetric cell divisions during animal development. *Developmental Cell*, 21(1), 102–119.
- Moujaber, O., & Stochaj, U. (2020). The cytoskeleton as regulator of cell signaling pathways. *Trends in Biochemical Sciences*, 45(2), 96–107.
- Murthy, K., & Wadsworth, P. (2005). Myosin-II-dependent localization and dynamics of F-actin during cytokinesis. *Current Biology*, 15(8), 724–731.
- Nakajima, Y. I. (2018). Mitotic spindle orientation in epithelial homeostasis and plasticity. *Journal of Biochemistry*, 164(4), 277–284.
- Otto, J. J., Kane, R. E., & Bryan, J. (1979). Formation of filopodia in coelomocytes: Localization of fascin, a 58,000 Dalton actin cross-linking protein. *Cell*, 17(2), 285–293.
- Quintero, O. A., & Yengo, C. M. (2012). Myosin X dimerization and its impact on cellular functions. *Proceedings of the National Academy of Sciences of the United States of America*, 109(43), 17313–17314.
- Rankin, E. B., & Giaccia, A. J. (2016). Hypoxic control of metastasis. *Science*, 352(6282), 175–180.

- Robinson, D. N., & Spudich, J. A. (2000). Towards a molecular understanding of cytokinesis. *Trends in Cell Biology*, 10(6), 228–237.
- Rogers, M. S., & Strehler, E. E. (2001). The tumor-sensitive calmodulin-like protein is a specific light chain of human unconventional myosin X. *The Journal of Biological Chemistry*, 276(15), 12182–12189.
- Sandquist, J. C., Larson, M. E., & Hine, K. J. (2016). Myosin-10 independently influences mitotic spindle structure and mitotic progression. *Cytoskeleton (Hoboken)*, 73(7), 351–364.
- Sato, O., Jung, H. S., Komatsu, S., Tsukasaki, Y., Watanabe, T. M., Homma, K., & Ikebe, M. (2017). Activated full-length myosin-X moves processively on filopodia with large steps toward diverse two-dimensional directions. *Scientific Reports*, 7, 44237.
- Savci-Heijink, C. D., Halfwerk, H., Koster, J., Horlings, H. M., & van de Vijver, M. J. (2019). A specific gene expression signature for visceral organ metastasis in breast cancer. *BMC Cancer*, 19(1), 333.
- Sigrist, C. J., De Castro, E., Langendijk-Genevaux, P. S., Le Saux, V., Bairoch, A., & Hulo, N. (2005). ProRule: A new database containing functional and structural information on PROSITE profiles. *Bioinformatics*, 21(21), 4060–4066.
- Sousa, A. D., & Cheney, R. E. (2005). Myosin-X: A molecular motor at the cell's fingertips. *Trends in Cell Biology*, 15(10), 533–539.
- Svitkina, T. (2018). The actin cytoskeleton and actin-based motility. *Cold Spring Harbor Perspectives in Biology*, 10(1), a018267. <http://doi.org/10.1101/cshperspect.a018267>.
- Tang, Z., Li, C., Kang, B., Gao, G., Li, C., & Zhang, Z. (2017). GEPIA: A web server for cancer and normal gene expression profiling and interactive analyses. *Nucleic Acids Research*, 45(W1), W98–W102.
- Thery, M., & Bornens, M. (2006). Cell shape and cell division. *Current Opinion in Cell Biology*, 18(6), 648–657.
- Titus, M. A. (2018). Myosin-driven intracellular transport. *Cold Spring Harbor Perspectives in Biology*, 10(3), a021972. <http://dx.doi.org/10.1101/cshperspect.a021972>.
- Tokuo, H., Bhawan, J., & Coluccio, L. M. (2018). Myosin X is required for efficient melanoblast migration and melanoma initiation and metastasis. *Scientific Reports*, 8(1), 10449.
- Tokuo, H., Mabuchi, K., & Ikebe, M. (2007). The motor activity of myosin-X promotes actin fiber convergence at the cell periphery to initiate filopodia formation. *The Journal of Cell Biology*, 179(2), 229–238.
- Torres, J. Z., Ban, K. H., & Jackson, P. K. (2010). A specific form of Phospho protein phosphatase 2 regulates anaphase-promoting complex/cyclosome association with spindle poles. *Molecular Biology of the Cell*, 21(6), 897–904.
- Torres, J. Z., Miller, J. J., & Jackson, P. K. (2009). High-throughput generation of tagged stable cell lines for proteomic analysis. *Proteomics*, 9(10), 2888–2891.
- Uraji, J., Scheffler, K., & Schuh, M. (2018). Functions of Actin in mouse oocytes at a glance. *Journal of Cell Science*, 131(22), jcs218099.
- Vasquez, C. G., Heissler, S. M., Billington, N., Sellers, J. R., Martin, A. C. (2016). Drosophila non-muscle myosin II motor activity determines the rate of tissue folding. *eLife*, 5. <http://dx.doi.org/10.7554/elife.20828>.
- Walczak, C. E., & Heald, R. (2008). Mechanisms of mitotic spindle assembly and function. *International Review of Cytology*, 265, 111–158.
- Weber, K. L., Sokac, A. M., Berg, J. S., Cheney, R. E., & Bement, W. M. (2004). A microtubule-binding myosin required for nuclear anchoring and spindle assembly. *Nature*, 431(7006), 325–329.
- Woolner, S., O'Brien, L. L., Wiese, C., & Bement, W. M. (2008). Myosin-10 and Actin filaments are essential for mitotic spindle function. *The Journal of Cell Biology*, 182(1), 77–88.
- Wuhr, M., Mitchison, T. J., & Field, C. M. (2008). Mitosis: New roles for myosin-X and actin at the spindle. *Current Biology*, 18(19), R912–R914.
- Yu, H., Chakravorty, S., Song, W., & Ferenczi, M. A. (2016). Phosphorylation of the regulatory light chain of myosin in striated muscle: Methodological perspectives. *European Biophysics Journal*, 45(8), 779–805.
- Zhang, L., Huang, S. T., Feng, Y. L., Wan, T., Gu, H. F., Xu, J., ... Zheng, M. (2017). The bidirectional regulation between MYL5 and HIF-1alpha promotes cervical carcinoma metastasis. *Theranostics*, 7(15), 3768–3780.
- Zhang, Y. S., Liu, B., Luo, X. J., Li, T. B., Zhang, J. J., Peng, J. J., ... Li, Q. (2015). Nuclear cardiac myosin light chain 2 modulates NADPH oxidase 2 expression in myocardium: A novel function beyond muscle contraction. *Basic Research in Cardiology*, 110(4), 38.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Ramirez I, Gholkar AA, Velasquez EF, et al. The myosin regulatory light chain Myl5 localizes to mitotic spindle poles and is required for proper cell division. *Cytoskeleton*. 2021;78:23–35. <https://doi.org/10.1002/cm.21654>

Regulation of Iron Homeostasis through Parkin-Mediated Lactoferrin Ubiquitylation

Ankur A. Gholkar, Stefan Schmollinger, Erick F. Velasquez, Yu-Chen Lo, Whitaker Cohn, Joseph Capri, Harish Dharmarajan, William J. Dearnorff, Lucy W. Gao, Mai Abdusamad, Julian P. Whitelegge, and Jorge Z. Torres*

Cite This: *Biochemistry* 2020, 59, 2916–2921

Read Online

ACCESS |

Metrics & More

Article Recommendations

Supporting Information

ABSTRACT: Somatic mutations that perturb Parkin ubiquitin ligase activity and the misregulation of iron homeostasis have both been linked to Parkinson's disease. Lactotransferrin (LTF) is a member of the family of transferrin iron binding proteins that regulate iron homeostasis, and increased levels of LTF and its receptor have been observed in neurodegenerative disorders like Parkinson's disease. Here, we report that Parkin binds to LTF and ubiquitylates LTF to influence iron homeostasis. Parkin-dependent ubiquitylation of LTF occurred most often on lysines (K) 182 and 649. Substitution of K182 or K649 with alanine (K182A or K649A, respectively) led to a decrease in the level of LTF ubiquitylation, and substitution at both sites led to a major decrease in the level of LTF ubiquitylation. Importantly, Parkin-mediated ubiquitylation of LTF was critical for regulating intracellular iron levels as overexpression of LTF ubiquitylation site point mutants (K649A or K182A/K649A) led to an increase in intracellular iron levels measured by ICP-MS/MS. Consistently, RNAi-mediated depletion of Parkin led to an increase in intracellular iron levels in contrast to overexpression of Parkin that led to a decrease in intracellular iron levels. Together, these results indicate that Parkin binds to and ubiquitylates LTF to regulate intracellular iron levels. These results expand our understanding of the cellular processes that are perturbed when Parkin activity is disrupted and more broadly the mechanisms that contribute to Parkinson's disease.

Parkinson's disease (PD) is a debilitating neurodegenerative disease whose incidence has increased over the past decade and presents a major public health epidemic.¹ The *PARK2* gene that encodes the Parkin E3 ubiquitin ligase is found to be mutated in familial forms of PD.² Two pathological hallmarks of PD are the aggregation of α -synuclein in Lewy bodies and the accumulation of iron.³ Lactotransferrin (lactoferrin or LTF) is a member of the family of transferrin iron binding proteins that transport iron and regulate intracellular iron levels.^{4–6} X-ray structures of LTF show that it displays a bilobal architecture with one iron binding site within each lobe.^{7,8} Increased levels of LTF and its receptor have been reported within nigral neurons in PD patients and in other neurodegenerative disorders like Alzheimer's disease.^{4,9–11} Iron homeostasis is important for maintaining the normal physiology of neuronal cell populations, and iron accumulation leads to neurotoxicity.¹² Due to this dysregulation of iron homeostasis during PD progression, iron chelators have been proposed as a potential therapeutic strategy.^{13,14} Although many models for how iron-mediated cell death occurs in PD exist, most agree that an excess of reactive iron (Fe^{2+}) leads to the generation of reactive oxygen species that induces oxidative stress and promotes neuronal cell death through toxic reactions that include mitochondrial dysfunction, lipid peroxidation, and protein misfolding.^{13,14} Interestingly, transferrin (Tf)-containing endosomes have been shown to contact the mitochondria and may function to deliver iron directly to the mitochondria,¹⁵ which could be a

mechanism that contributes to mitochondrial dysfunction. However, whether other members of the transferrin family, like LTF, function in a similar manner remains unclear. Although the dysregulation of Parkin and LTF had been previously linked to the accumulation of iron in populations of neuronal cells that undergo cell death in PD, a direct connection between the two had not been established.

To better understand the role of Parkin dysfunction in PD, we sought to identify novel Parkin ubiquitylation substrates. First, we established a HEK293 doxycycline-inducible localization and affinity purification (LAP = EGFP-TEV-S-Peptide)-tagged Parkin stable cell line and utilized it to express and tandem affinity-purified LAP-Parkin.^{16,17} Eluates were analyzed by mass spectrometry to identify Parkin-associated proteins (Figure 1A). This analysis identified Parkin (270 peptides) and lactotransferrin (LTF, 72 peptides) as the most abundant proteins (Figure 1B and Table S1). Additionally, tubulin isoforms and proteasome subunits, known Parkin-interacting proteins,^{18–23} were also identified along with subunits of the CCT/TRiC complex (chaperonin-containing T-complex/TCP-1 ring complex), which is critical for tubulin

Received: June 16, 2020
 Revised: July 22, 2020
 Published: August 4, 2020



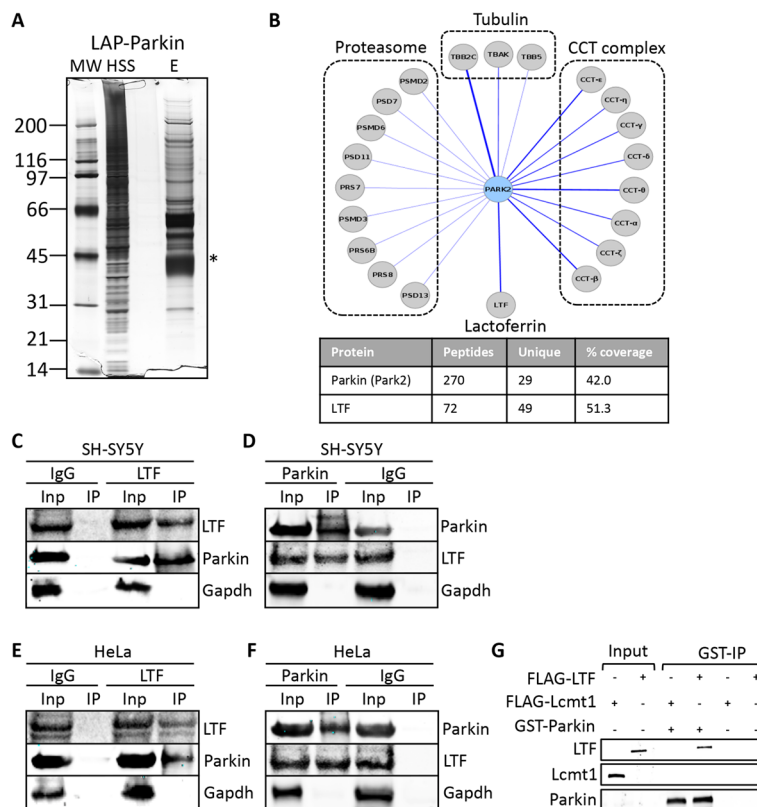


Figure 1. LTF co-purifies and associates with Parkin. (A) LAP–Parkin tandem affinity purification. Abbreviations: MW, molecular weight marker; HSS, high-spin supernatant; E, final eluates. The asterisk denotes the Parkin protein band. (B) Cytoscape visualization map of Parkin-associated proteins, identified by mass spectrometry, showing the major classes of co-purifying proteins. The bottom panel highlights the identification of lactoferrin (LTF) as a Parkin co-purifying protein. The protein name, the number of peptides identified, the number of unique peptides, and the percent protein coverage are indicated. See Table S1 for a complete list of Parkin co-purifying proteins identified by mass spectrometry. (C–F) SH-SY5Y or HeLa cell extracts were used to perform reciprocal co-immunoprecipitation (Co-IP) experiments using anti-lactoferrin (LTF), anti-Parkin, and control IgG antibodies. Note that endogenous Parkin immunoprecipitates with endogenous LTF (C and E) and endogenous LTF immunoprecipitates with endogenous Parkin (D and F). Inp indicates input, and IP indicates immunoprecipitation. (G) *In vitro* binding assays performed in the presence or absence of *in vitro*-transcribed/translated FLAG–LTF, FLAG–Lcmt1, or GST–Parkin. GST–Parkin was immunoprecipitated (GST-IP), and eluates were analyzed by immunoblotting with the indicated antibodies. Note that FLAG–LTF immunoprecipitates with GST–Parkin, whereas control FLAG–Lcmt1 does not.

folding and for blocking the fibrillation of α -synuclein that is a pathological hallmark of PD (Figure 1B and Table S1).^{24–26} Due to the importance of LTF in iron homeostasis and its misregulation in PD,⁴ we sought to further validate the Parkin–LTF interaction. Reciprocal co-immunoprecipitation experiments from SH-SY5Y neuronal cells and HeLa cells with anti-Parkin and anti-LTF antibodies showed that LTF co-immunoprecipitated with Parkin and Parkin co-immunoprecipitated with LTF (Figure 1C–F). Similarly, *in vitro* protein binding reactions with GST–Parkin and FLAG–LTF showed that LTF co-immunoprecipitated with Parkin (Figure 1G). Together, these data indicated that LTF was associating with Parkin.

To understand the significance of Parkin–LTF association, we asked if LTF was ubiquitylated and whether its ubiquitylation was Parkin-dependent. LAP–LTF was immunoprecipitated from control siRNA (siCont) or Parkin siRNA (siParkin)-treated cells, and its ubiquitylation was monitored by immunoblot analysis with anti-ubiquitin antibodies. LAP–LTF was ubiquitylated in the siCont sample, and the level of this ubiquitylation was substantially decreased upon Parkin depletion with siParkin (Figure 2A). Next, we asked if LTF was a Parkin substrate using an *in vitro*-reconstituted

ubiquitylation assay.²⁷ GST–LTF, GST–tubulin (positive control), or GST–GFP (negative control) were incubated with an ATP regeneration system, ubiquitin, an E1 ubiquitin-activating enzyme, an E2 ubiquitin-conjugating enzyme, and wild-type (WT) or LAP–Parkin-overexpressing HEK293 cell extracts. Parkin substrate ubiquitylation was then monitored by immunoprecipitating the GST-tagged proteins and performing an immunoblot analysis with anti-ubiquitin and anti-GST antibodies. We observed ubiquitylation of GST–LTF and GST–tubulin (a known substrate of Parkin¹⁸) as a ladder of increasing molecular weight bands (Figure 2B). Next, we analyzed LTF ubiquitylation reactions by mass spectrometry and determined that LTF was ubiquitylated at seven different lysine (K) residues, with K182 and K649 being the most frequently modified sites (Table S2). Mapping of the ubiquitylation sites onto the human LTF crystal structure [Protein Data Bank (PDB) entry 1FCK] showed that all sites were on exposed loops (Figure 2C).

Next, we analyzed the contribution of the most frequently modified lysines (K182 or K649) to the overall ubiquitylation of LTF by substituting them with alanines and assessing LTF ubiquitylation. LAP–LTF-WT or LTF single (K182A or K649A) or double (K182A/K649A) ubiquitylation site point

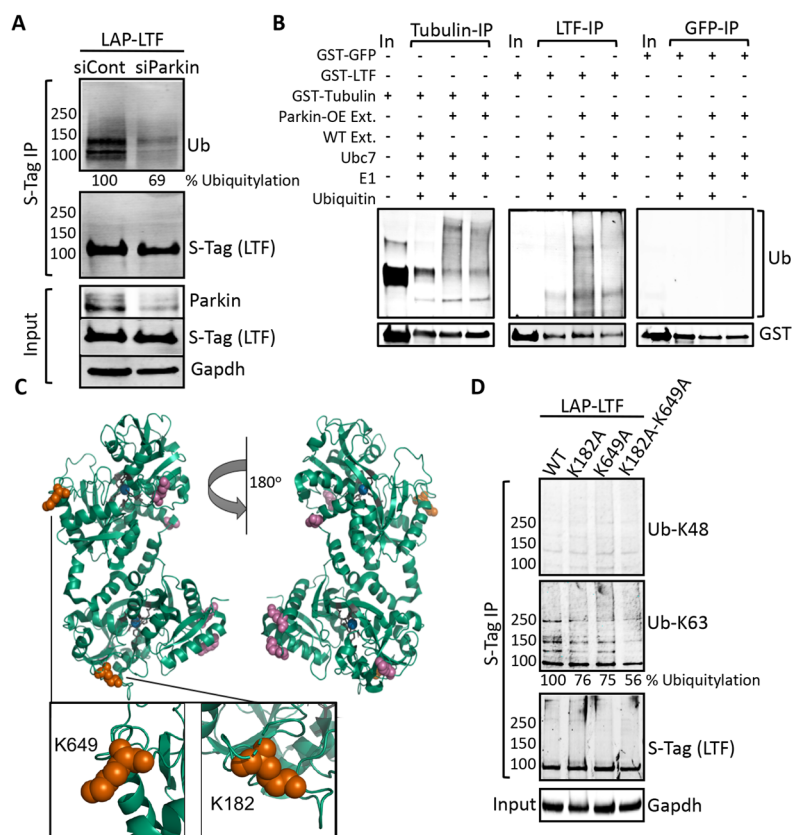


Figure 2. LTF is a substrate of Parkin. (A) Immunoblot analysis of LAP–LTF immunoprecipitated from HeLa cells treated with control nontargeting siRNA (siCont) or siRNA targeting Parkin (siParkin). Immunoprecipitation of LAP–LTF (S-Tag-IP) was monitored with anti-S-tag antibodies. LAP–LTF ubiquitylation was monitored with anti-ubiquitin antibodies (Ub). Depletion of Parkin was monitored with anti-Parkin antibodies, and anti-Gapdh antibodies were used as a control to monitor Gapdh protein levels. IP indicates immunoprecipitation. (B) *In vitro* ubiquitylation assays with or without recombinant GST–LTF, GST–tubulin (positive control), or GST–GFP (negative control); an ATP regeneration system; ubiquitin; E1 ubiquitin-activating enzyme; E2 ubiquitin-conjugating enzyme (Ubc7); and wild-type (WT) or LAP–Parkin-overexpressing (OE) HEK293 cell extracts (Ext). GST-tagged proteins were immunoprecipitated, and their ubiquitylation was monitored with anti-ubiquitin antibodies (Ub). IP indicates immunoprecipitation. (C) LTF ubiquitylation reactions were analyzed by mass spectrometry, and the most abundant Parkin-mediated LTF ubiquitylation sites, K182 and K649, were mapped onto the human LTF structure of PDB entry 1FCK (represented as orange spheres). Five additional ubiquitylation sites are represented as magenta spheres. Fe bound to LTF is represented as blue spheres. For a complete list of identified LTF ubiquitylation sites, see Table S2. (D) LAP–LTF-WT or LTF single (K182A or K649A) or double (K182A/K649A) ubiquitylation site mutants were expressed in HeLa cells and immunoprecipitated (S-Tag IP), and their ubiquitylation was monitored using anti-K48 and anti-K63 ubiquitin linkage specific antibodies. IP indicates immunoprecipitation.

mutants were expressed in HeLa cells and immunoprecipitated, and their ubiquitylation status was monitored using anti-K48 and anti-K63 ubiquitin linkage specific antibodies. K182A or K649A single mutants showed a reduction in the level of K63-linked ubiquitylation of LTF, and ubiquitylation of the LTF K182A/K649A double mutant was highly impaired compared to the WT control (Figure 2D). Together, these data indicated that LTF was a Parkin substrate and that K182A and K649A were the most frequently ubiquitylated sites and accounted for the majority of LTF-ubiquitylated species. Interestingly, LTF protein levels remained unchanged in HeLa or SH-SY5Y cells treated with Parkin siRNA or in Parkin knockout mice brains (Figure S1), indicating that Parkin did not regulate the levels of LTF.

Next, we sought to determine if ubiquitylation at K182 or K649 could influence the ability of LTF to regulate intracellular iron levels. Extracts from control HeLa cells or HeLa cells overexpressing the LTF wild type, K182A single mutant, K649A single mutant, or K182A/K649A double mutant were analyzed for intracellular sulfur (S), iron (Fe),

and zinc (Zn) levels using inductively coupled plasma mass spectrometry (ICP-MS/MS). For all ICP-MS/MS analyses, Fe and Zn levels were normalized to total S levels. Overexpression of the K649A variant alone or the K182A/K649A double mutant resulted in a similar significant increase in the intracellular Fe level compared to the control, while the Zn levels did not significantly change (Figure 3A,B). We hypothesized that if Parkin was ubiquitylating LTF on K649A to regulate iron levels, then modulation of Parkin levels would also affect intracellular iron levels. To test this, we performed RNAi-mediated depletion of Parkin levels and again analyzed the extracts for total S, Fe, and Zn. Decreasing Parkin levels led to a significant increase in Fe content compared to the control, while the Zn levels did not significantly change (Figure 4A,B). In contrast, extracts from cells that were overexpressing Parkin showed a significant decrease in total Fe levels compared to the control, while the Zn levels did not significantly change (Figure 4C,D). Together, these data demonstrated that Parkin abundance directly or the substitution of the Parkin-dependent LTF ubiquitylation site

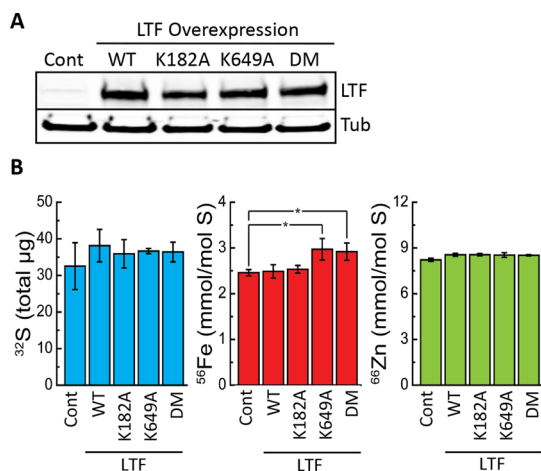


Figure 3. Overexpression of LTF ubiquitylation site mutants influences intracellular iron levels. (A) Extracts from control HeLa cells (Cont) or HeLa cells overexpressing the LAP–LTF wild type (WT), lysine 182 to alanine (K182A) point mutant, lysine 649 to alanine (K649A) point mutant, or double point mutant (K182A/K649A, indicated by DM) were analyzed for sulfur (³²S), iron (⁵⁶Fe), and zinc (⁶⁶Zn). (B) Total ³²S levels (in micrograms) were used to quantify ⁵⁶Fe and ⁶⁶Zn levels (in millimoles per mole of S). Bar graphs show the mean ± standard deviation from three replicate samples. A *t* test was used to calculate *p* values ($\alpha < 5\%$) in the indicated comparisons.

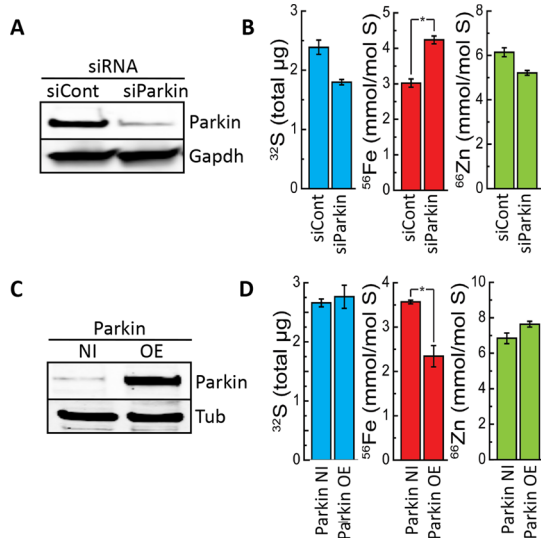


Figure 4. Parkin levels influence intracellular iron levels. (A and B) Extracts from HeLa cells transfected with control nontargeting siRNA (siCont) or Parkin siRNA (siParkin) were analyzed for total sulfur (³²S), iron (⁵⁶Fe), and zinc (⁶⁶Zn). (C and D) Extracts from the HEK293 LAP–Parkin-inducible cell line that was not induced (Parkin NI) or induced to overexpress LAP–Parkin (Parkin OE) were analyzed for total sulfur (³²S), iron (⁵⁶Fe), and zinc (⁶⁶Zn). NI indicates not induced, and OE indicates overexpression. (B and D) The total ³²S levels (in micrograms) were used to quantify ⁵⁶Fe and ⁶⁶Zn levels (in millimoles per mole of S). Bar graphs show the mean ± standard deviation from three replicate samples. A *t* test was used to calculate *p* values ($\alpha < 5\%$) in the indicated comparisons.

(K649A) affects intracellular iron levels, consistent with a role for Parkin in regulating iron levels through LTF ubiquitylation. To begin to understand how Parkin-mediated ubiquitylation of LTF on K649 was affecting LTF metal binding, we performed

molecular dynamics simulations to compare the interaction potential energy in the LTF structure to the monoubiquitylated LTF structure. LTF structural studies have shown that the structure of LTF remains essentially the same irrespective of whether iron (Fe), cerium (Ce), or copper (Cu) is bound.^{28–31} For this analysis, we used the structure of PDB entry 1FCK that is bound to Ce.³⁰ First, the monoubiquitylated LTF was modeled by conjugating the ubiquitin structure C-terminal glycine to K649 of human lactoferrin (see **Materials and Methods**) (Figure S2A). The molecular dynamics simulation showed that the monoubiquitylated LTF had an overall Ce binding stability that was lower than that of LTF with a mean initial interaction energy of approximately -2120 kcal/mol compared to a value of -2170 kcal/mol. Furthermore, the stability of monoubiquitylated LTF decreased over the 100 ps simulation with an increased interaction energy while the Ce metal binding for LTF remained stable during the simulation (Figure S2B). Inspection of energy-minimized LTF and monoubiquitylated LTF structures revealed a shift in residues His597, Tyr435, and Arg465 that coordinate metal binding.²⁹ This *in silico* molecular analysis indicated that LTF ubiquitylation at K649 led to structural changes that destabilize metal coordination in the binding site and therefore could modulate metal (including iron) binding activity.

In summary, we have discovered a previously undescribed link between Parkin and LTF that influences iron homeostasis. We identified lactoferrin (LTF or lactotransferrin) as a Parkin-interacting protein and validated this interaction through reciprocal co-immunoprecipitations from cells and in binding reactions *in vitro*. Increased levels of LTF and its receptor have been reported within nigral neurons in PD patients and in other neurodegenerative disorders.^{4,9} Iron homeostasis is important for maintaining normal physiology of neuronal cell populations, and iron accumulation leads to neurotoxicity.¹² Due to the importance of LTF in iron homeostasis and its misregulation in PD, we sought to further define the significance of the Parkin–LTF interaction. We determined that LTF was ubiquitylated in a Parkin-dependent manner in cells and *in vitro* through K63 linkages. Moreover, we mapped the sites of seven lysines on the LTF lysine that were ubiquitylated, with K182 and K649 being the most abundantly modified. Substitution of K182 or K649 with an alanine (K182A or K649A) led to a decrease in the level of LTF ubiquitylation, and the double point mutant led to a major decrease in the level of LTF ubiquitylation. Importantly, Parkin-mediated ubiquitylation of LTF was critical for LTF’s ability to modulate iron levels as overexpression of LTF ubiquitylation site point mutants K649A or K182A/K649A led to an increase in intracellular iron levels measured by ICP-MS/MS. Consistently, RNAi-mediated depletion of Parkin also led to an increase in intracellular iron levels in contrast to overexpression of Parkin that led to a decrease in intracellular iron levels. Together, our data suggest that Parkin binds to and ubiquitylates LTF to influence intracellular iron levels. We propose that Parkin ubiquitylation of LTF at K649 perturbs LTF’s ability to accumulate intracellular iron levels and that depletion of Parkin, or substitution of K649 on LTF, allows LTF to accumulate intracellular iron levels. The ability of Parkin to influence iron levels through LTF ubiquitylation may have direct implications for the increased iron levels that are observed in the nigral cells of PD patients. These results expand our understanding of the cellular processes that are

perturbed when Parkin activity is disrupted and more broadly the mechanisms that contribute to PD. LTF is a member of the family of transferrin iron binding proteins that transport iron and regulate intracellular iron levels.^{4,5} It is currently unclear whether other isoforms of LTF like δ -lactoferrin³² or other transferrin iron binding proteins are regulated through ubiquitylation and whether their ubiquitylation is Parkin-mediated. This study should help stimulate additional work related to Parkin-mediated iron regulation in PD and other neurological disorders. We note that the regulation of iron levels through Parkin-mediated ubiquitylation of LTF that we observe *in vitro* cannot solely account for the dysregulation in iron levels seen in PD patients and that other mechanisms are likely to exist. For example, the protein levels of divalent metal transporter 1 (DMT1), which transports metals, including iron, into the cell from the extracellular environment, have been shown to be regulated through Parkin-mediated ubiquitylation and proteasome-dependent degradation, and elevated levels of DMT1 have been observed in PD patients.^{4,33,34} Furthermore, it is important to note that LTF can also influence the levels of other proteins with roles in iron homeostasis.³⁵ For example, studies analyzing the effect of bovine lactoferrin (bLf) on the inflammation response showed that the addition of exogenous bLf could inhibit iron overload by reducing the level of production of the pro-inflammatory cytokines IL-6 and IL-1 β and regulating iron metabolism through the upregulation of the iron exporter ferroportin (Fpn) and the transferrin receptor 1 (TfR1) and down-regulation of ferritin (Ftn).^{35,36} Thus, whether the changes in intracellular iron levels that we observe upon LTF ubiquitylation are directly related to LTF function or to the function of downstream iron-regulating proteins that are modulated by LTF remains to be determined.

■ ASSOCIATED CONTENT

SI Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.biochem.0c00504>.

Tables, figures, and experimental methods (PDF)

Accession Codes

LTF, PO2788; Park2, 060260.

■ AUTHOR INFORMATION

Corresponding Author

Jorge Z. Torres – Department of Chemistry and Biochemistry, Jonsson Comprehensive Cancer Center, and Molecular Biology Institute, University of California, Los Angeles, California 90095, United States; orcid.org/0000-0002-2158-889X; Phone: 310-206-2092; Email: torres@chem.ucla.edu

Authors

Ankur A. Gholkar – Department of Chemistry and Biochemistry, University of California, Los Angeles, California 90095, United States

Stefan Schmollinger – Department of Chemistry and Biochemistry and Institute for Genomics and Proteomics, University of California, Los Angeles, California 90095, United States

Erick F. Velasquez – Department of Chemistry and Biochemistry, University of California, Los Angeles, California 90095, United States

Yu-Chen Lo – Department of Chemistry and Biochemistry, University of California, Los Angeles, California 90095, United States

Whitaker Cohn – Pasarow Mass Spectrometry Laboratory, University of California, Los Angeles, California 90095, United States

Joseph Capri – Pasarow Mass Spectrometry Laboratory, University of California, Los Angeles, California 90095, United States

Harish Dharmarajan – Department of Chemistry and Biochemistry, University of California, Los Angeles, California 90095, United States

William J. Deardorff – Department of Chemistry and Biochemistry, University of California, Los Angeles, California 90095, United States

Lucy W. Gao – Pasarow Mass Spectrometry Laboratory, University of California, Los Angeles, California 90095, United States

Mai Abdusamad – Department of Chemistry and Biochemistry, University of California, Los Angeles, California 90095, United States

Julian P. Whitelegge – Pasarow Mass Spectrometry Laboratory, Jonsson Comprehensive Cancer Center, and Molecular Biology Institute, University of California, Los Angeles, California 90095, United States; orcid.org/0000-0003-2763-7733

Complete contact information is available at:

<https://pubs.acs.org/10.1021/acs.biochem.0c00504>

Author Contributions

A.A.G. and J.Z.T. contributed to the design, execution, and analysis of experiments. M.A., H.D., and W.J.D. contributed to the experimentation. S.S. performed the ICP-MS/MS analyses. E.F.V. and Y.-C.L. performed LTF structure modeling. W.C., J.C., L.W.G., and J.P.W. performed mass spectrometry characterization of LTF ubiquitylation. A.A.G. and J.Z.T. wrote the final version of the manuscript.

Funding

This work was supported by National Science Foundation Grant NSF-MCB1912837 to J.Z.T., UCSD/UCLA NIDDK Diabetes Research Center Grant P30 DK063491 to J.P.W., U.S. Department of Energy Grant DE-FD02-04ER15529 to support S.S., a UCLA Molecular Biology Institute Whitcome Fellowship to E.F.V., and National Institutes of Health Grant GM42143 for instrument support.

Notes

The authors declare no competing financial interest.

■ ABBREVIATIONS

PD, Parkinson's disease; PARK2, Parkin; LTF, lactotransferrin; LAP, localization and affinity purification; ICP-MS/MS, inductively coupled plasma mass spectrometry.

■ REFERENCES

- (1) Connolly, B. S., and Lang, A. E. (2014) Pharmacological treatment of Parkinson disease: a review. *JAMA* 311, 1670–1683.
- (2) Panicker, N., Dawson, V. L., and Dawson, T. M. (2017) Activation mechanisms of the E3 ubiquitin ligase parkin. *Biochem. J.* 474, 3075–3086.
- (3) Lingor, P., Carboni, E., and Koch, J. C. (2017) Alpha-synuclein and iron: two keys unlocking Parkinson's disease. *J. Neural Transm (Vienna)* 124, 973–981.
- (4) Hirsch, E. C. (2009) Iron transport in Parkinson's disease. *Parkinsonism and Related Disorders* 15 (Suppl. 3), S209–S211.

- (5) Mazurier, J., Legrand, D., Hu, W. L., Montreuil, J., and Spik, G. (1989) Expression of human lactotransferrin receptors in phytohemagglutinin-stimulated human peripheral blood lymphocytes. Isolation of the receptors by antiligand-affinity chromatography. *Eur. J. Biochem.* 179, 481–487.
- (6) Qian, Z. M., and Wang, Q. (1998) Expression of iron transport proteins and excessive iron accumulation in the brain in neurodegenerative disorders. *Brain Res. Rev.* 27, 257–267.
- (7) Thomassen, E. A., van Veen, H. A., van Berkel, P. H., Nuijens, J. H., and Abrahams, J. P. (2005) The protein structure of recombinant human lactoferrin produced in the milk of transgenic cows closely matches the structure of human milk-derived lactoferrin. *Transgenic Res.* 14, 397–405.
- (8) Andersen, B. F., Baker, H. M., Morris, G. E., Rumball, S. V., and Baker, E. N. (1990) Apolactoferrin structure demonstrates ligand-induced conformational change in transferrins. *Nature* 344, 784–787.
- (9) Faucheux, B. A., Herrero, M. T., Villares, J., Levy, R., Javoy-Agid, F., Obeso, J. A., Hauw, J. J., Agid, Y., and Hirsch, E. C. (1995) Autoradiographic localization and density of [¹²⁵I]ferrotransferrin binding sites in the basal ganglia of control subjects, patients with Parkinson's disease and MPTP-lesioned monkeys. *Brain Res.* 691, 115–124.
- (10) Leveugle, B., Faucheux, B. A., Bouras, C., Nillesse, N., Spik, G., Hirsch, E. C., Agid, Y., and Hof, P. R. (1996) Cellular distribution of the iron-binding protein lactotransferrin in the mesencephalon of Parkinson's disease cases. *Acta Neuropathol.* 91, 566–572.
- (11) Kawamata, T., Tooyama, I., Yamada, T., Walker, D. G., and McGeer, P. L. (1993) Lactotransferrin immunocytochemistry in Alzheimer and normal human brain. *Am. J. Pathol.* 142, 1574–1585.
- (12) Ward, R. J., Zucca, F. A., Duyn, J. H., Crichton, R. R., and Zecca, L. (2014) The role of iron in brain ageing and neurodegenerative disorders. *Lancet Neurol.* 13, 1045–1060.
- (13) Weinreb, O., Mandel, S., Youdim, M. B. H., and Amit, T. (2013) Targeting dysregulation of brain iron homeostasis in Parkinson's disease by iron chelators. *Free Radical Biol. Med.* 62, 52–64.
- (14) Nunez, M. T., and Chana-Cuevas, P. (2019) New perspectives in iron chelation therapy for the treatment of Parkinson's disease. *Neural Regener. Res.* 14, 1905–1906.
- (15) Sheftel, A. D., Zhang, A. S., Brown, C., Shirihai, O. S., and Ponka, P. (2007) Direct interorganellar transfer of iron from endosome to mitochondrion. *Blood* 110, 125–132.
- (16) Torres, J. Z., Miller, J. J., and Jackson, P. K. (2009) High-throughput generation of tagged stable cell lines for proteomic analysis. *Proteomics* 9, 2888–2891.
- (17) Bradley, M., Ramirez, I., Cheung, K., Gholkar, A. A., and Torres, J. Z. (2016) Inducible LAP-tagged Stable Cell Lines for Investigating Protein Function, Spatiotemporal Localization and Protein Interaction Networks. *J. Visualized Exp.* 118, 54870.
- (18) Ren, Y., Zhao, J., and Feng, J. (2003) Parkin binds to alpha/beta tubulin and increases their ubiquitination and degradation. *J. Neurosci.* 23, 3316–3324.
- (19) Collins, G. A., and Goldberg, A. L. (2020) Proteins containing ubiquitin-like (Ubl) domains not only bind to 26S proteasomes but also induce their activation. *Proc. Natl. Acad. Sci. U. S. A.* 117, 4664–4674.
- (20) Upadhyaya, S. C., and Hegde, A. N. (2003) A potential proteasome-interacting motif within the ubiquitin-like domain of parkin and other proteins. *Trends Biochem. Sci.* 28, 280–283.
- (21) Sakata, E., Yamaguchi, Y., Kurimoto, E., Kikuchi, J., Yokoyama, S., Yamada, S., Kawahara, H., Yokosawa, H., Hattori, N., Mizuno, Y., Tanaka, K., and Kato, K. (2003) Parkin binds the Rpn10 subunit of 26S proteasomes through its ubiquitin-like domain. *EMBO Rep.* 4, 301–306.
- (22) Aguilera, M. A., Korac, J., Durcan, T. M., Trempe, J. F., Haber, M., Gehring, K., Elsasser, S., Waidmann, O., Fon, E. A., and Husnjak, K. (2015) The E3 ubiquitin ligase parkin is recruited to the 26 S proteasome via the proteasomal ubiquitin receptor Rpn13. *J. Biol. Chem.* 290, 7492–7505.
- (23) Safadi, S. S., and Shaw, G. S. (2010) Differential interaction of the E3 ligase parkin with the proteasomal subunit S5a and the endocytic protein Eps15. *J. Biol. Chem.* 285, 1424–1434.
- (24) Lundin, V. F., Leroux, M. R., and Stirling, P. C. (2010) Quality control of cytoskeletal proteins and human disease. *Trends Biochem. Sci.* 35, 288–297.
- (25) Sot, B., Rubio-Munoz, A., Leal-Quintero, A., Martinez-Sabando, J., Marcilla, M., Roodvelde, C., and Valpuesta, J. M. (2017) The chaperonin CCT inhibits assembly of alpha-synuclein amyloid fibrils by a specific, conformation-dependent interaction. *Sci. Rep.* 7, 40859.
- (26) Swinnen, E., Buttner, S., Outeiro, T. F., Galas, M. C., Madeo, F., Winderickx, J., and Franssens, V. (2011) Aggresome formation and segregation of inclusions influence toxicity of alpha-synuclein and synphilin-1 in yeast. *Biochem. Soc. Trans.* 39, 1476–1481.
- (27) Laney, J. D., and Hochstrasser, M. (2011) Analysis of protein ubiquitination. *Curr. Protoc. Protein Sci.* 66, 14.15.11–14.15.13.
- (28) Anderson, B. F., Baker, H. M., Dodson, E. J., Norris, G. E., Rumball, S. V., Waters, J. M., and Baker, E. N. (1987) Structure of human lactoferrin at 3.2-Å resolution. *Proc. Natl. Acad. Sci. U. S. A.* 84, 1769–1773.
- (29) Baker, E. N., Anderson, B. F., Baker, H. M., Haridas, M., Jameson, G. B., Norris, G. E., Rumball, S. V., and Smith, C. A. (1991) Structure, function and flexibility of human lactoferrin. *Int. J. Biol. Macromol.* 13, 122–129.
- (30) Baker, H. M., Baker, C. J., Smith, C. A., and Baker, E. N. (2000) Metal substitution in transferrins: specific binding of cerium(IV) revealed by the crystal structure of cerium-substituted human lactoferrin. *JBIC, J. Biol. Inorg. Chem.* 5, 692–698.
- (31) Smith, C. A., Anderson, B. F., Baker, H. M., and Baker, E. N. (1992) Metal substitution in transferrins: the crystal structure of human copper-lactoferrin at 2.1-Å resolution. *Biochemistry* 31, 4527–4533.
- (32) Siebert, P. D., and Huang, B. C. (1997) Identification of an alternative form of human lactoferrin mRNA that is expressed differentially in normal tissues and tumor-derived cell lines. *Proc. Natl. Acad. Sci. U. S. A.* 94, 2198–2203.
- (33) Roth, J. A., Singleton, S., Feng, J., Garrick, M., and Paradkar, P. N. (2010) Parkin regulates metal transport via proteasomal degradation of the 1B isoforms of divalent metal transporter 1. *J. Neurochem.* 113, 454–464.
- (34) Salazar, J., Mena, N., Hunot, S., Prigent, A., Alvarez-Fischer, D., Arredondo, M., Duyckaerts, C., Sazdovitch, V., Zhao, L., Garrick, L. M., Nunez, M. T., Garrick, M. D., Raisman-Vozari, R., and Hirsch, E. C. (2008) Divalent metal transporter 1 (DMT1) contributes to neurodegeneration in animal models of Parkinson's disease. *Proc. Natl. Acad. Sci. U. S. A.* 105, 18578–18583.
- (35) Rosa, L., Cutone, A., Lepanto, M. S., Paesano, R., and Valenti, P. (2017) Lactoferrin: A Natural Glycoprotein Involved in Iron and Inflammatory Homeostasis. *Int. J. Mol. Sci.* 18, 1985.
- (36) Cutone, A., Rosa, L., Lepanto, M. S., Scotti, M. J., Berlutti, F., Bonaccorsi di Patti, M. C., Musci, G., and Valenti, P. (2017) Lactoferrin Efficiently Counteracts the Inflammation-Induced Changes of the Iron Homeostasis System in Macrophages. *Front. Immunol.* 8, 705.