Exploring variation in protein half lives across tissues and inferring stability control mechanisms through causal inference

Andrew Leduc, Devon Kohler

Protein abundance is a function of both synthesis and degradation. However, the extent to which variation in degradation rates across proteins and cell types determines variation of protein levels is an open question. Anecdotally, there are a number of examples illustrating the control of protein abundances through degradation such as the regulation of cell cycle progression [1] and the response to hypoxia [1,2] by the ubiquitin proteasome system (UPS). Specific recognition of target proteins by E3 ubiquitin ligases provides a mechanism for protein specific degradation control, and variation in the activity of these ligases provides mechanisms for differential degradation for a given protein across conditions. The existence of over 600 distinct E3 ubiquitin ligases suggests that tuning of degradation rates may control protein abundance more broadly across the proteome [3].

Indeed, the distribution of protein half lives across the proteome has been reported to span two orders of magnitude from roughly 1 hours to 1000 hours both in vitro [4,5] and in vivo [6,7]. Additionally, these studies report significant deviations in half-lives across cell types [4] and tissues [6]. Despite compelling anecdotal examples and significant variation in half lives, systematic explorations of the relationship between protein levels across different proteins have reported only a modest inverse correlation to protein half-life, with estimates of contributed variance being as low as 5% [5,8]. However, the relationship between protein abundance and half-lives remains relatively unexplored as studies have made measurements in few *in vitro* systems [5,7] or achieved limited depth of protein coverage[6].

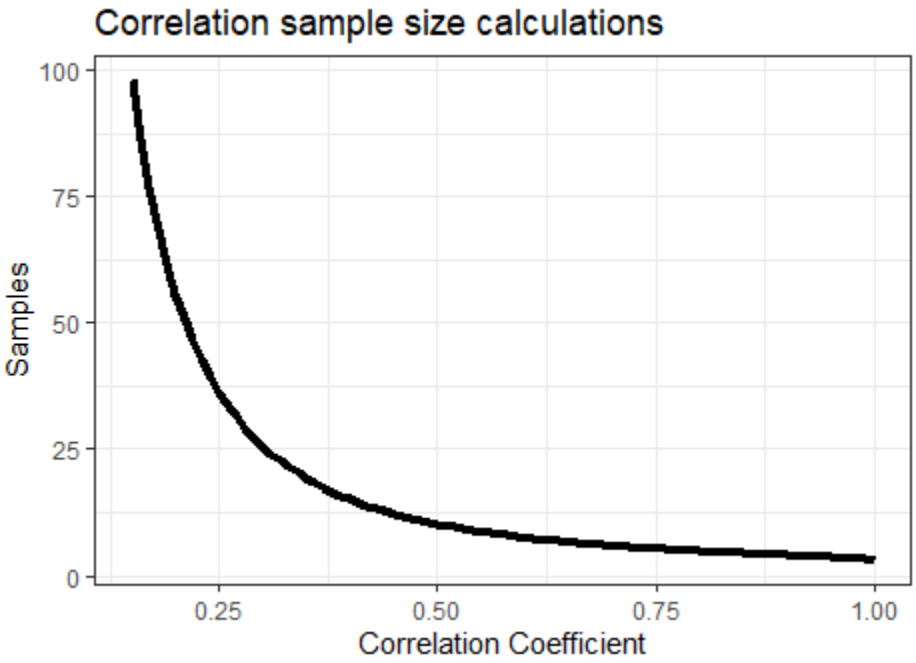
Recent advances in protein sequencing via liquid chromatography coupled with tandem mass spectrometry (LC-MS/MS) have made it possible to quantify thousands of proteins from small tissue samples, even single cells. This opens new possibilities and significantly reduces the required mass spectrometry analysis time, and thus cost and the amount of labor required to prepare samples. To improve our understanding of how variation in protein half-lives shapes variation in protein abundance, we propose leveraging these advances to measure protein half-lives at proteome wide scale across a dozen murine tissues spanning lifespan and genders. This can be achieved through *in vivo* metabolic pulse experiments, where mice have their diet changed to feed containing isotopically heavy lysine and arginine isotopes. The relative abundances of light proteins synthesized before the diet change and heavy proteins synthesized after the diet change can then be used to infer protein half lives across all measured proteins, while the abundance of proteins can be inferred from the total amount of heavy and light proteins. Our preliminary data and analysis suggests that 1) there is significant variation in protein half lives across different biological contexts and 2) protein degradation rates can be a major determinant of protein abundance, especially in cell types and tissues with low cell growth and proliferation rates.

Further, we propose to unpack the influence of protein half life on protein abundance through modeling the relationship between the various proteins that control degradation rate and protein half life. We plan on modeling the relationships as a complex network of related interactions. There are many modeling methods that can be employed to infer the underlying regulatory network. One straightforward method is using conditional independence tests. These tests can determine if the network structure implied by the graph is in agreement with the experimental data. While effective, running conditional independence tests on all possible relationships between proteins is prohibitively computationally expensive. We propose leveraging prior knowledge databases, such as INDRA [10,11], to create a prior network of relationships, limiting the search space. We will then use the prior network and the acquired data to estimate a posterior network across different biological contexts.

Budget Report

Item	Cost	Rationale
1kg food	$15,000	Minimum amount of food available for purchase
Mice	$5,000	Sufficient replication is required to test correlations between proteins. We calculate the replicates below given a power of .8 and a FDR of .05.[12]
Total	$20,000	

Sample size calculations for correlation analysis:



References:

1. Hegde, A. N. & van Leeuwen, F. W. *Ubiquitin and the Brain: Roles of Proteolysis in the Normal and Abnormal Nervous System*. (Frontiers Media SA, 2017).
2. Salceda, S. & Caro, J. Hypoxia-inducible factor 1alpha (HIF-1alpha) protein is rapidly degraded by the ubiquitin-proteasome system under normoxic conditions. Its stabilization by hypoxia depends on redox-induced changes. *J. Biol. Chem.* **272**, 22642–22647 (1997).
3. Yang, Q., Zhao, J., Chen, D. & Wang, Y. E3 ubiquitin ligases: styles, structures and functions. *Mol Biomed* **2**, 23 (2021).
4. Mathieson, T. *et al.* Systematic analysis of protein turnover in primary cells. *Nat. Commun.* **9**, 689 (2018).
5. Schwanhäusser, B. *et al.* Global quantification of mammalian gene expression control. *Nature* **473**, (2011).
6. Rolfs, Z. *et al.* An atlas of protein turnover rates in mouse tissues. *Nat. Commun.* **12**, 1–9 (2021).
7. Fornasiero, E. F. *et al.* Precisely measured protein lifetimes in the mouse brain reveal differences across tissues and subcellular fractions. *Nat. Commun.* **9**, 1–17 (2018).
8. Schoenheimer, R. *The Dynamic State of Body Constituents*. (1949).
9. Kohler, D. *et al.* MSstats Version 4.0: Statistical Analyses of Quantitative Mass Spectrometry-Based Proteomic Experiments with Chromatography-Based Quantification at Scale. *J. Proteome Res.* (2023) doi:10.1021/acs.jproteome.2c00834.
10. Gyori, B.M. *et al*. From word models to executable models of signaling networks using automated assembly (2017), Molecular Systems Biology, 13, 954.
11. Bachman, J.A. *et al*. Automated assembly of molecular mechanisms at scale from text mining and curated databases (2023), Molecular Systems Biology, e11325.
12. Hulley, SB *et al*. Designing clinical research : an epidemiologic approach. 4th ed. Philadelphia, PA: Lippincott Williams & Wilkins; 2013. Appendix 6C, page 79.