

EMBARGOED until Feb 27th 2014: 7 am EST; 12 midday UK time

System-wide analyses have underestimated protein abundances and the importance of transcription in animals.

Over the last ten years, a number of studies have suggested that, in animal cells, translation and protein turnover play a larger role in determining the different levels at which proteins are expressed than transcription. The major evidence supporting these claims is a weak correlation between system-wide protein and mRNA abundance measurements. A highly cited Nature article by Schwanhausser et al. in 2011 provides the most comprehensive example of such analyses. A new study just published in PeerJ by Li et al., however, questions the conclusions of these papers. This new study suggests that the major reason why protein and mRNA abundance measurements are poorly correlated is because of various types of measurement error in the protein and mRNA abundance, rather than transcription having minimal impact on protein expression levels.

Li et al. first show that Schwanhausser et al.'s protein abundances have a non linear error that leads to a dramatic underestimation of low abundance proteins, a result that has been independently supported by a separate benchmarking study by Ahrne et al. Li et al. rescale Schwanhausser et al.'s protein abundance estimates using data for housekeeping proteins and show that the rescaled data show a higher correlation with mRNA abundances than the uncorrected protein data. In addition, they estimate the impact of other sources of error on the mRNA and protein abundance measurements using direct experimental data, and they find that, when error is explicitly measured and modeled, an even greater correlation between mRNA and protein is expected. Li et al. use a second, independent strategy to determine the contribution of mRNA levels to protein expression: they show that the variance in translation rates directly measured by ribosome profiling is dramatically lower than that inferred by Schwanhausser et al., and that the measured and inferred translation rates correlate poorly. Incorporating protein and mRNA turnover data in this analysis, the results from Li et al. suggest that mRNA levels explain ~81% of the variance in protein levels, transcription 71%, RNA degradation 10%; translation 11%; and protein degradation 8%. This conclusion differs dramatically from the previous estimates of differences in mRNA levels explaining 10-40% of the differences in protein levels in the current literature.

Li et al.'s analysis provides an accurate framework for quantifying gene expression and protein abundance levels by explicitly considering sources of error. This work highlights the importance of appropriate statistical analyses of the large quantitative data sets that are increasingly being produced by experimentalists and are being used to study fundamental cellular mechanisms.

###

EMBARGOED until February 27th 2014: 7 am EST; 12 midday UK time

Link to the PDF of this Press Release: <http://bit.ly/LiPeerJRelease>

Link to the Press Preview of the Original Article (this link should only be used **BEFORE** the embargo ends): <http://static.peerj.com/press/previews/2014/02/270.pdf> (note: this is an author proof and so may change slightly before publication)

Link to the Published Version of the article (quote this link in your story – the link will **ONLY** work **after** the embargo lifts): <https://peerj.com/articles/270> - your readers will be able to **freely** access this article at this URL.

Citation to the article: Li JJ, Bickel PJ, Biggin MD. (2014) System wide analyses have underestimated protein abundances and the importance of transcription in mammals. *PeerJ* 2:e270
<http://dx.doi.org/10.7717/peerj.270>

Funding: This work was supported in part by NIH grant P01 GM009655. Work at Lawrence Berkeley National Laboratory was conducted under Department of Energy contract DEAC02-05CH11231

###

About PeerJ

PeerJ is an Open Access publisher of peer reviewed articles, which offers researchers a lifetime publication plan, for a single low price, providing them with the ability to openly publish all future articles for free. PeerJ is based in San Francisco, CA and London, UK and can be accessed at <https://peerj.com/>.

All works published in PeerJ are Open Access and published using a Creative Commons license (CC-BY 4.0). Everything is immediately available—to read, download, redistribute, include in databases and otherwise use—without cost to anyone, anywhere, subject only to the condition that the original authors and source are properly attributed.

PeerJ Media Resources (including logos) can be found at: <https://peerj.com/about/press/>

###

Media Contacts

Note: If you would like to join the PeerJ Press Release list, visit: <http://bit.ly/PressList>

For the Authors: Mark Biggin, Tel: +1 510 486 7606, Email: mdbiggin@lbl.gov

For PeerJ: Email: press@peerj.com ; <https://peerj.com/about/press/>

###

Abstract (from the article)

Large scale surveys in mammalian tissue culture cells suggest that the protein expressed at the median abundance is present at 8,000 - 16,000 molecules per cell and that differences in mRNA expression between genes explain only 10-40% of the differences in protein levels. We find, however, that these surveys have significantly underestimated protein abundances and the relative importance of transcription. Using individual measurements for 61 housekeeping proteins to rescale whole proteome data from Schwanhausser et al., we find that the median protein detected is expressed at 170,000 molecules per cell and that our corrected protein abundance estimates show a higher correlation with mRNA abundances than do the uncorrected protein data. In addition, we estimated the impact of further errors in mRNA and protein abundances using direct experimental measurements of these errors. The resulting analysis suggests that mRNA levels explain at least 56% of the differences in protein abundance for the 4,212 genes detected by Schwanhausser et al., though because one major source of error could not be estimated the true percent contribution should be higher. We also employed a second, independent strategy to determine the contribution of mRNA levels to protein expression. We show that the variance in translation rates directly measured by ribosome profiling is only 12% of that inferred by Schwanhausser et al. and that the measured and inferred translation rates correlate poorly ($R^2=0.13$). Based on this, our second strategy suggests that mRNA levels explain ~81% of the variance in protein levels. We also determine the percent contributions of transcription, RNA degradation, translation and protein degradation to the variance in protein abundances using both of our strategies. While the magnitudes of the two estimates vary, they both suggest that transcription plays a more important role than the earlier studies implied and translation a much smaller role. Finally, the above estimates only apply to those genes whose mRNA and protein expression was detected. Based on a detailed analysis by Hebenstreit et al., we estimate that approximately 40% of genes in a given cell within a population express no mRNA. Since there can be no translation in the absence of mRNA, we argue that differences in translation rates can play no role in determining the expression levels for the ~40% of genes that are non-expressed.