

successful. These pioneering initiatives are beginning to link pharmaceutical companies, academia, and disease experts across the “gene to bedside” spectrum in the locations where these diseases have the greatest impact. The MMV provides free, open access to a range of compounds with activity against a range of pathogens for independent researchers to screen, with users requested to publish their data in the public domain, thus continuing the drug development research cycle (17).

The impact of bacterial AMR in low-income countries is severe and likely to worsen. New antimicrobial agents may provide some respite against AMR and infections caused by such drug-resistant pathogens. However, introducing novel broad-range antimicrobials into the current melee of antimicrobial use and misuse in lower-income countries would only have a short-term limited impact on infections caused by potentially life-threatening pathogens. Restricting the use of the same classes of antimicrobial compounds in animals and humans has to be an immediate priority, including a direct ban of any new antimicrobials developed for treating infections in humans. Lastly, new antimicrobial agents should only be administered to those who really need them. This means that the current capacity to perform microbial diagnostics and downstream antimicrobial susceptibility testing needs to be greatly improved, alongside the development of rational prescribing practice. ■

REFERENCES AND NOTES

1. Department of Health & Department for Environment Food & Rural Affairs. *UK Five Year Antimicrobial Resistance Strategy 2013 to 2018* (2013).
2. Centers for Disease Control and Prevention, Antibiotic Resistance Threats in the United States, 2013; www.cdc.gov/drugresistance/threat-report-2013/pdf/ar-threats-2013-508.pdf.
3. World Health Organization. “Antimicrobial resistance: Global report on surveillance” (2014).
4. M. Perros. *Science* **347**, 1062 (2015).
5. T. R. Walsh *et al.*, *Lancet Infect. Dis.* **11**, 355 (2011).
6. H. Chang The *et al.*, *EMBO Mol. Med.* **10**, 15252/emmm.201404767 (2015).
7. S. P. Stone *et al.*, *BMJ* **344**, e3005 (2012).
8. K. E. Holt *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **110**, 17522 (2013).
9. A. Mutreja *et al.*, *Nature* **477**, 462 (2011).
10. P. Roumagnac *et al.*, *Science* **314**, 1301 (2006).
11. S. Kariuki *et al.*, *J. Clin. Microbiol.* **48**, 2171 (2010).
12. C. M. Parry *et al.*, *PLOS Negl. Trop. Dis.* **5**, e1163 (2011).
13. K. D. Koirala *et al.*, *Antimicrob. Agents Chemother.* **56**, 2761 (2012).
14. M. N. Robertson *et al.*, *Parasitology* **141**, 148 (2014).
15. See www.gatesfoundation.org/What-We-Do/Global-Health/Tuberculosis.
16. See www.mmv.org.
17. K. Ingram-Sieber *et al.*, *PLOS Negl. Trop. Dis.* **8**, e2610 (2014).

ACKNOWLEDGMENTS

S. B. is a Sir Henry Dale Fellow, jointly funded by the Wellcome Trust and the Royal Society (100087/Z/12/Z).

10.1126/science.aaa2868

GENE EXPRESSION

Statistics requantitates the central dogma

Transcription, not translation, chiefly determines protein abundance in mammals

By Jingyi Jessica Li¹ and Mark D. Biggin²

Mammalian proteins are expressed at $\sim 10^3$ to 10^8 molecules per cell (1). Differences between cell types, between normal and disease states, and between individuals are largely defined by changes in the abundance of proteins, which are in turn determined by rates of transcription, messenger RNA (mRNA) degradation, translation, and protein degradation. If the rates for one of these steps differ much more than the rates of the other three, that step would be dominant in defining the variation in protein expression. Over the past decade, system-wide studies have claimed that in animals, differences in translation rates predominate (2–5). On page 1112 of this issue, Jovanovic *et al.* (6), as well as recent studies by Battle *et al.* (7) and Li *et al.* (1), challenge this conclusion, suggesting that transcriptional control makes the larger contribution.

Earlier studies used mass spectrometry, DNA microarrays, and mRNA sequencing (mRNA-Seq) to measure protein and mRNA levels for thousands of genes (2–5), and also to measure the rates of mRNA degradation, translation, and/or protein degradation (through labeling with stable isotopes) (4, 5). Some studies examined a single cell type at steady state (5), whereas others analyzed the differences between tissue types (4), between tumors (2), or between inbred mouse strains (3). Each study found a moderate to low correlation between protein and mRNA abundance data (coefficient of determination $R^2 \leq 0.4$). This was taken to suggest that no more than 40% of the variance in protein levels is explained by variance in the rates of transcription and mRNA degradation and, by implication, that the remaining variance in protein expression ($\geq 60\%$) is explained by translation and protein degradation (2–5). By employing degradation rate data for mRNAs and proteins in addition to abundance data, it was further estimated that transcription explains 34% of the variance in protein abundance, mRNA degradation 6%, translation 55%, and protein degrada-

tion 5% (5) (see the figure).

The high-throughput methods used in these studies, however, show substantial stochastic variation between replica data and also suffer systematic, reproducible biases (1, 4–6, 8, 9). For example, label-free mass spectrometry can underestimate amounts of all lower-abundance proteins by as much as a factor of 10 (1, 8), and mRNA-Seq data are biased by guanine-cytosine base pair content by a factor of up to 3 (9). Because each type of error has different causes and because RNA and protein techniques differ greatly, the errors should be uncorrelated. Thus, the correlation of protein versus mRNA as measured will be lower than that between error-free data. The papers by Jovanovic *et al.*, Battle *et al.*, and Li *et al.* used careful statistical efforts to estimate and/or reduce the impact of errors and thereby find the higher correlation expected between true protein and true mRNA levels.

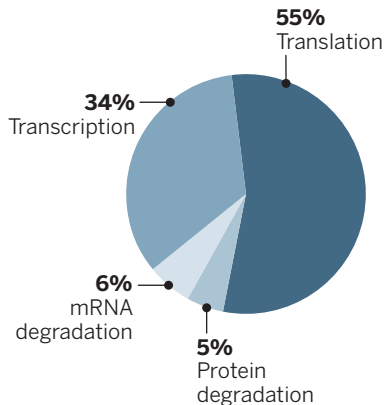
“Before gene expression can be correctly modeled, an accurate accounting of molecular abundances and expression rate constants is vital.”

Jovanovic *et al.* examined mouse bone marrow dendritic cells at steady state and during response to bacterial lipopolysaccharide (LPS) (6). They used a Bayesian model to estimate the true rates of translation and protein destruction from noisy mass spectrometry data. In addition, three independent estimates of protein abundance were made from three samples, each digested with a different protease. These three differently biased estimates were then used in separate parts of the analy-

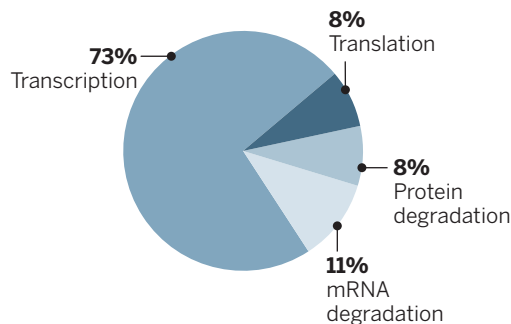
¹Department of Statistics and Department of Human Genetics, University of California, Los Angeles, CA 90095, USA. ²Genomics Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA. E-mail: jjli@stat.ucla.edu; mdbiggin@lbl.gov

% Contribution to protein levels

Original data estimates



Error-corrected estimates



Control of protein expression. The charts show the percent contributions of the variance in the rates of each step in gene expression to the variance in protein abundance for 4212 genes (from a mouse cell line). The left chart shows estimates from (5); the right chart shows estimates from (1) that take into account stochastic and systematic errors in the abundance data of (5).

sis to avoid a confounding dependency on common errors that would result if a single estimate were used throughout. Filtering of unreliable data and estimates of stochastic mRNA-Seq errors, in addition, allowed Jovanovic *et al.* to calculate that at steady state, mRNA levels explain 68% of the variance in protein expression, translation rates 26%, and protein degradation rates 8%. Upon stimulation of cells with LPS, mRNA levels appear to explain 90% of the changes in protein expression, with translation and protein degradation explaining only 4% and 6%, respectively. Jovanovic *et al.* did find, though, that upon LPS treatment, translation and protein degradation rates changed more for ribosomal, mitochondrial, and other highly expressed housekeeping proteins than for other genes, indicating an important role for these two steps in the control of some processes.

Battle *et al.* took a different tack, examining human protein variation among 62 individuals from the Yoruba population of Ibadan, Nigeria (7). Genomic DNA sequences for each individual were compared to mRNA-Seq, ribosome footprinting (ribosome density per mRNA), and mass spectrometry data for lymphoblastoid cells derived from each person. Consistent with previous results (3), the variation in measured protein levels between individuals correlates poorly with the variation in measured mRNA abundances (mean $R^2 < 0.2$) (7). However, when only those differences in expression that are associated with variation in the DNA sequence of a nearby gene were considered, most gene loci showing

changes in protein levels between individuals also showed correlated differences in mRNA expression, consistent with a dominant role for transcription. In addition, there was “a scarcity” of DNA sequence changes that affected only ribosome footprint density and protein abundance, not mRNA levels. In effect, by constraining their analysis to only those differences in expression associated with DNA sequence variation, Battle *et al.* excluded much of the variation due to measurement errors to obtain a more accurate answer.

Li *et al.* (1) (our own study) reanalyzed data in (5) with two approaches to account for measurement errors. In the first, a nonlinear scaling error in protein abundance estimates (from mass spectrometry data) was corrected using classic data from the literature, and a subset of the other errors in the mRNA-Seq and protein abundance data was estimated from replica and other control data. In the second approach, variance in translation rates measured directly by ribosome footprinting was substituted for a larger variance that had been inferred indirectly with a model in (5). The first approach suggests that the variance in true mRNA levels explains a minimum of 56% of the variance in true protein levels. The second implies that true mRNA levels explain 84% of the variance in true protein expression, transcription 73%, RNA degradation 11%, and translation and protein degradation each only 8%.

Most controllers of gene expression identified by classic genetic or biochemical methods are either transcription factors or proteins (such as kinases and signaling

receptors) that directly regulate the activities of proteins, not their abundances. In addition, translation and mRNA degradation rates change only modestly upon cellular differentiation or when microRNA expression is perturbed (10–12). Moreover, improved statistical analyses show that in contrast to earlier studies, mRNA levels explain most of the variance in protein abundances in yeast (13, 14). Finally, ~40% of genes in a single mammalian cell express no mRNA (1, 15); thus, for these ~8800 genes, transcriptional repression by chromatin is likely the sole determinant of the absence of protein expression.

Understanding the contributions of transcriptional versus posttranscriptional control is not simply a matter of academic interest. For example, variation in protein expression among 95 colorectal tumor samples is only poorly explained by measured mRNA abundances (2), which might imply that different responses of patients to anticancer treatments are posttranscriptional effects. If, however, most of the variation in protein levels is controlled by transcription but this fact is obscured by measurement errors, then differences in drug action could be mainly explained by variation at the transcriptional level.

Accurate quantitation of the control of gene expression is in its infancy. Experimental protocols with fewer inherent biases are needed, along with further improvements in statistical methods that can estimate and take error into account. Before gene expression can be correctly modeled, an accurate accounting of molecular abundances and expression rate constants is vital. ■

REFERENCES AND NOTES

1. J. J. Li, P. J. Bickel, M. D. Biggin, *PeerJ* **2**, e270 (2014).
2. B. Zhang *et al.*, *Nature* **513**, 382 (2014).
3. A. Ghazalpour *et al.*, *PLoS Genet.* **7**, e1001393 (2011).
4. A. R. Kristensen, J. Gsponer, L. J. Foster, *Mol. Syst. Biol.* **9**, 689 (2013).
5. B. Schwanhäusser *et al.*, *Nature* **473**, 337 (2011).
6. M. Jovanovic *et al.*, *Science* **347**, 1259038 (2015).
7. A. Battle *et al.*, *Science* **347**, 664 (2015).
8. E. Ahmri, L. Molzahn, T. Glatter, A. Schmidt, *Proteomics* **13**, 2567 (2013).
9. M. S. Cheung, T. A. Down, I. Latorre, J. Ahinger, *Nucleic Acids Res.* **39**, e103 (2011).
10. D. Baek *et al.*, *Nature* **455**, 64 (2008).
11. M. Selbach *et al.*, *Nature* **455**, 58 (2008).
12. N. T. Ingolia, L. F. Lareau, J. S. Weissman, *Cell* **147**, 789 (2011).
13. G. Csardi *et al.*, <http://biorxiv.org/content/early/2014/12/26/009472> (2014).
14. F. W. Albert, D. Muzzey, J. S. Weissman, L. Kruglyak, *PLoS Genet.* **10**, e1004692 (2014).
15. D. Hebenstreit, A. Deonarine, M. M. Babu, S. A. Teichmann, *Curr. Opin. Cell Biol.* **24**, 350 (2012).

ACKNOWLEDGMENTS: J.J.L. was supported in part by the Department of Statistics at UCLA. Work at Lawrence Berkeley Laboratory National Laboratory was conducted under U.S. Department of Energy contract DEAC02-05CH11231.

10.1126/science.aaa8332

This copy is for your personal, non-commercial use only.

If you wish to distribute this article to others, you can order high-quality copies for your colleagues, clients, or customers by [clicking here](#).

Permission to republish or repurpose articles or portions of articles can be obtained by following the guidelines [here](#).

The following resources related to this article are available online at www.sciencemag.org (this information is current as of March 5, 2015):

Updated information and services, including high-resolution figures, can be found in the online version of this article at:

<http://www.sciencemag.org/content/347/6226/1066.full.html>

A list of selected additional articles on the Science Web sites **related to this article** can be found at:

<http://www.sciencemag.org/content/347/6226/1066.full.html#related>

This article **cites 14 articles**, 2 of which can be accessed free:

<http://www.sciencemag.org/content/347/6226/1066.full.html#ref-list-1>

This article appears in the following **subject collections**:

Immunology

<http://www.sciencemag.org/cgi/collection/immunology>