measured in PBS[4], which may result in an overestimation of photostability compared to commonly used live-cell imaging conditions. The use of media depleted of vitamins for fluorescence imaging of live cultured cells appears to be a simple and efficient way to improve the performance of some widely used fluorescent proteins in various ensemble and single-molecule applications[1,5,6].

Note: Supplementary information is available on the Nature Methods website.

ACKNOWLEDGMENTS

This work was supported by Russian Academy of Sciences (Molecular and Cell Biology program and Innovation and Development support 31-236), Howard Hughes Medical Institute (55005618) and Rosnauka grants 02.512.11.2216 and 02.512.12.2053. D.M.C. and K.A.L. are supported by President of the Russian Federation grants MK-6119.2008.4 and MD-2780.2009.4.

COMPETING INTERESTS STATEMENT

The authors declare competing financial interests: details accompany the full-text HTML version of the paper at http://www.nature.com/naturemethods/.

Alexey M Bogdanov[1], Ekaterina A Bogdanova[1], Dmitriy M Chudakov[1], Tatiana V Gorodnicheva[2], Sergey Lukyanov[1] & Konstantin A Lukyanov[1]

[1]Shemiakin-Ovchinnikov Institute of Bioorganic Chemistry, Moscow, Russia.
[2]Evrogen JSC, Moscow, Russia.
e-mail: kluk@ibch.ru

1. Shaner, N.C., Patterson, G.H. & Davidson, M.W. *J. Cell Sci.* **120**, 4247–4260 (2007).
2. Bogdanov, A.M. *et al. Nat. Chem. Biol.* **5**, 459–461 (2009).
3. Werner, R., Manthey, K.C., Griffin, J.B. & Zempleni, J. *J. Nutr. Biochem.* **16**, 617–624 (2005).
4. Shaner, N.C., Steinbach, P.A. & Tsien, R.Y. *Nat. Methods* **2**, 905–909 (2005).
5. Kohl, T. & Schwille, P. *Adv. Biochem. Eng. Biotechnol.* **95**, 107–142 (2005).
6. Fernandez-Suarez, M. & Ting, A.Y. *Nat. Rev. Mol. Cell. Biol.* **9**, 929–943 (2008).

Recurated protein interaction datasets

To the Editor: In their recent Perspective, Cusick *et al.*[1] state "…curation can be error-prone and possibly of lower quality than commonly assumed." Although we welcome rigorous scrutiny of curation efforts, Cusick *et al.*[1] had arrived at their conclusions by misunderstanding the difference between the reliability of experimental data supporting protein interactions and the correctness of the curation process itself.

The aim of the IntAct molecular interaction database (IntAct)[2], the Database of Interacting Proteins (DIP)[3], the Molecular Interaction database (MINT)[4], the *Arabidopsis* Information Resource (TAIR)[5] and the Biological General Repository for Interaction Datasets (BioGRID)[6] critiqued by Cusick *et al.*[1] is to collect and organize experiments supporting protein-protein interactions into a comprehensive set of accurately annotated experimental data. These databases allow the biological community facile and searchable access to a vast repository of biological interactions for many purposes ranging from individual hypothesis generation to functional annotation to biological network analysis. The transparent and full representation of interactions in the primary literature is an essential component of such a repository and is necessary to assess the reliability of published data. As databases support many different uses of their data, they aim to incorporate the complete data as presented in the source publications, rather than selecting evidence they consider more reliable or otherwise privileged. The use of detailed and well-defined controlled vocabularies for annotation allows the database users to efficiently select subsets of data according to criteria relevant for their particular use.

In contrast, Cusick *et al.*[1] define a set of criteria for a specific use restricted only to direct pairwise protein-protein interactions, which they refer to as 'binary' interactions. They evaluate literature-curated datasets against these criteria and then assert that failure to meet their criteria represents "incorrect curation." The criteria defined by Cusick *et al.*[1] vary slightly from species to species but aim to select only direct interactions with multiple independent supporting reports. While this is one valid use, other users might, for example, look for all observed interactions of a given protein, whether direct or indirect, to subsequently assess the supporting evidence by reading the supporting publications. Whereas protein-protein interaction databases may also use the term 'binary' when referring to pairs of interacting proteins, our usage of the term refers to any interaction pair and makes no judgment regarding whether the interaction is direct or indirect.

We strongly object to the notion that inclusion of an interaction with limited supporting evidence of a direct interaction represents a curation error. On the contrary, most interaction databases always fully curate a given publication and would consider it an egregious omission if only a subset of the protein interactions reported in a publication or its supplementary material would be contained in the database. When information—for example, species information—in a publication is ambiguous, database curators attempt to contact the authors and only leave out data if clarification cannot be obtained.

In response to the claims of Cusick *et al.*[1], we reanalyzed interactions presented in their paper to identify actual curation errors, defined as inconsistencies between the original published data and their representation in our databases. Details of our analysis are available in the **Supplementary Note**, and we reannotated versions of the original tables supplied by Cusick *et al.*[1] (**Supplementary Tables 1–3**). The actual curation error rate was, in fact, consistently under 10%.

For the yeast dataset, we confirmed 4 actual curation errors among the 100 sample interactions from BioGRID chosen by Cusick *et al.*[1]; the curation error rate of 4% is precisely the value originally reported for the dataset[7] and an order of magnitude lower than the claim by Cusick *et al.*[1]: "Of the interacting pairs in the sample, 35% were incorrectly curated." For comparison, we analyzed a subset of the BioGRID data that is also present in the DIP database and identified 1 actual curation error out of 29 shared records, that is, a similarly low error rate of 3%.

For the human dataset, of the 220 sampled interactions annotated in MINT, only 10 were curation errors, corresponding to a curation error rate of 4.5%. Similarly, only 4 out of 42 curation records reported in DIP contained errors, a 9% curation error rate, or one-fifth of the 45% curation error rate implied by Cusick *et al.*[1].

For the *Arabidopsis thaliana*, the IntAct dataset contained 3 actual curation errors in 183 curation records, resulting in an error rate of 2%, less than one-fifth of the 10.7% rate claimed by Cusick *et al.*[1] in their Table 2. For TAIR, the actual error rate was only 3%, or less than one-third of the rate claimed by Cusick *et al.*[1].

Accurate and detailed curation is an arduous process both in terms of individual curator expertise and curation time. To optimize the use of public funding, the member databases of the International Molecular Exchange Consortium (IMEx)[8] DIP, IntAct and MINT coordinate their curation efforts to avoid unnecessary redundancy,

as described on the consortium webpage (http://imex.sf.net/). The low overlap between IMEx interaction datasets noted by Cusick *et al.*[1] is not, as claimed, an indicator for undersampling of the interaction space, but rather demonstrates the success of the international collaboration within the IMEx consortium.

In summary, when appropriately considering only actual curation errors rather than subjective reliability criteria intended to identify only the subset of directly interacting protein pairs, our analysis demonstrated a surprisingly narrow spread of 2–9% curation errors across datasets from three different species curated by five different interaction databases. This analysis testified to the precision of interaction database curation and substantiated the case for coordinated international efforts to curate biological interactions.

Note: Supplementary information is available on the Nature Methods website.

Lukasz Salwinski[1,10], Luana Licata[2,10], Andrew Winter[3,10], David Thorneycroft[4], Jyoti Khadake[4], Arnaud Ceol[2], Andrew Chatr Aryamontri[2], Rose Oughtred[5], Michael Livstone[5], Lorrie Boucher[6], David Botstein[5], Kara Dolinski[5], Tanya Berardini[7], Eva Huala[7], Mike Tyers[3,6], David Eisenberg[1,8], Gianni Cesareni[2,9] & Henning Hermjakob[4]

[1]University of California, Los Angeles Department of Energy Institute for Genomics and Proteomics, Los Angeles, California, USA. [2]Instituto di Ricovero e Cura a Carattere Scientificio Fondazione Santa Lucia, Rome, Italy. [3]School of Biological Sciences, The University of Edinburgh, Edinburgh, UK. [4]European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, UK. [5]Lewis-Sigler Institute for Integrative Genomics, Princeton University, Princeton, New Jersey, USA. [6]Samuel Lunenfeld Research Institute, Mount Sinai Hospital, Toronto, Ontario, Canada. [7]Department of Plant Biology, Carnegie Institution for Science, Stanford, California, USA. [8]Department of Chemistry and Biochemistry, Howard Hughes Medical Institute, University of California, Los Angeles, California, USA. [9]Department of Biology, University of Rome Tor Vergata, Rome, Italy. [10]These authors contributed equally to this work.

e-mail: hhe@ebi.ac.uk, cesareni@uniroma2.it, david@mbi.ucla.edu or m.tyers@ed.ac.uk

1. Cusick, M.E. *et al. Nat. Methods* **6**, 39–46 (2009).
2. Kerrien, S. *et al. Nucleic Acids Res.* **35**, D561–D565 (2007).
3. Salwinski, L. *et al. Nucleic Acids Res.* **32**, D449–D451 (2004).
4. Chatr-aryamontri, A. *et al. Nucleic Acids Res.* **35**, D572–D574 (2007).
5. Swarbreck, D. *et al. Nucleic Acids Res.* **36**, D1009–D1014 (2008).
6. Breitkreutz, B.J. *et al. Nucleic Acids Res.* **36**, D637–D640 (2008).
7. Reguly, T. *et al. J. Biol.* **5**, 11 (2006).
8. Orchard, S. *et al. Proteomics* **7** (Suppl. 1), 28–34 (2007).

*Editor's note: For the response by Cusick et al., please see the Addendum to their Perspective (Cusick, M.E. et al. Nat. Methods **6**, 934–935; 2009).*