Transparency issues that undermine reproducibility

- **Methods descriptions**
  Common issues: randomization, blinding, sample size determinations, independent experiments vs technical replicates

- **Statistical clarity**
  - Gels, microscopy images unduly manipulated
  - Missing controls and markers
  - Reagent validity – antibodies, cell lines
  - Animal studies description
  - Data access, deposition & reusability

Needs joint approach from funders, publishers, institutions, researchers
Methodological details

Reporting checklists
## Reporting checklist of statistical and methodological details

Please ensure that the answers to the following questions are reported in the manuscript itself. We encourage you to include a specific subsection in the Methods section for statistics, reagents and animal models. **Below, provide the page numbers (s) for where the information can be located.**

<table>
<thead>
<tr>
<th>Statistics and General Methods</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. How was the sample size chosen to ensure adequate power to detect a pre-specified effect size? (Give section/paragraph or page number)</td>
</tr>
<tr>
<td>For animal studies, include a statement about sample size estimate even if no statistical methods were used. <em>For example, “No statistical method was used to predetermine sample size.”</em></td>
</tr>
<tr>
<td>2. Describe inclusion/exclusion criteria if samples or animals were excluded from the analysis. Were the criteria pre-established? (Give section/paragraph or page number)</td>
</tr>
</tbody>
</table>

Reproducibility checklist also currently being trialled at various BMC journals, including *BMC Biology, BMC Neuroscience, Genome Biology,* and *GigaScience.*
Raising reporting standards for data description

Checklist to improve figure legends and reporting

(a) Western blot of cell lysates of control and Rac1-siRNA-treated MTLn3 cells, blotted for Rac1 and β-actin. A representative image is shown from 3 blots. (b) MTLn3 cells transfected with control or Rac1 siRNA and plated on Alexa-405-conjugated gelatin overnight. Arrows point to invadopodia and sites of degradation. Scale bars, 10 μm. Representative image sets are shown from 50 image sets each for the control and Rac1 siRNA. (c) Quantification of mean degradation area per cell from b, including Rac1 inhibitor NSC23766 treatment at 100 μM. \( n = 60 \) fields for each condition, pooled from 5 independent experiments; error bars are s.e.m. Student’s \( t \)-test was used. **\( P = 0.00022, ^{\wedge}^{\wedge} P = 0.011639 \). Uncropped images of blots are shown in Supplementary Fig. 9.
Maximizing the reuse of research techniques
Protocol publication adds value and provides credit
A peer-reviewed journal of laboratory protocols. Each is presented in a 'recipe' style, with step-by-step descriptions which users can immediately apply in their own research.

Open resource maintained by NPG: collaborative, open (CC-NC), free to post and read
Access to data underlying the paper
Transparency for data
Fundamental sharing policy for *Nature* and the Nature research journals

An inherent principle of publication is that others should be able to replicate and build upon the authors' published claims. A condition of publication in a Nature journal is that *authors are required to make materials, data, code, and associated protocols promptly available* to readers without undue qualifications. Any restrictions on the availability of materials or information must be disclosed to the editors ... [and] ... in the submitted manuscript.

Supporting data must be made available to editors and peer-reviewers at the time of submission for the purposes of evaluating the manuscript.

See http://www.nature.com/authors/policies/availability.html
Step one: removing barriers to sharing

- Nature-titles and *Scientific Data* explicitly allow pre-publication sharing of data and article preprints.
- Publication of data articles will not compromise novelty of subsequent research articles.
- Similar policies at the BMC journals.
Solution for small datasets

Source data – aka “data behind the graph”

- Non-processed data and numerical values that went into making the figure
- Presented in a file that can be downloaded from the figure legend

(a) Expression of GCaMP3.0 in the mushroom body of an individual fly, focused on the horizontal lobes. The red line indicates a region of interest used to determine changes in fluorescence emission. Scale bar represents 50 μm. (b) False color-coded image of Ca²⁺ activity in the horizontal mushroom body lobes shown in a. Warm colors indicate high levels, cold colors indicate low levels or no Ca²⁺ activity. The numbers indicate changes in fluorescence ΔF/F₀ (%). (c,d) Time course of Ca²⁺ increase in horizontal mushroom body lobes of 3-d-old and 30-d-old Sdp⁻ flies evoked by the odorants 4-methyl cyclohexanol (MCH) or 3-octanol (OCT) in comparison with the diluent, mineral oil (n = 5 flies, non-parametric Mann-Whitney U test found no substantial difference between the Ca²⁺ increase of 3-d-old and 30-d-old Sdp⁻ flies evoked by the odorants). The gray bars indicate the duration of the odor stimuli. Data are presented as mean ± s.e.m.
Solution for small datasets

Source data – aka “data behind the graph”

- Non-processed data and numerical values that went into making the figure
- Presented in a file that can be downloaded from the figure legend

Allows to plot traces for individual flies from fig 3c with OCT treatment
Data-access practices strengthened in Nature journals

*Nature* editorial (Nov 2014)

- Clear preference for sharing large datasets via public repositories.
- Enforce data deposition in fields where there is strong community consensus.
- List of public data repositories now maintained by *Scientific Data*.
- Encourage authors to publish Data Descriptors at *Scientific Data*.
  - before, with or after the analysis paper.
  - editors work with authors.
Maximizing the reuse of data
Data journals add value and provide credit for reusable data
Data publication

Get Credit for Sharing Data
Publications will be listed in the major indexes and will be citeable.

Focused on Data Reuse
All the information others need to reuse the data; no interpretative analysis or hypothesis testing.

Open-access
The main article is published by default under the CC BY licence. Each publication supported by curated CC0 metadata.

Peer-reviewed
Rigorous peer-review managed by our Editorial Board of academic researchers ensures data quality and standards.

Data Note article-type available at two BMC journals.

Promoting Community Data Repositories
Data stored in community data repositories.
Global integrated drought monitoring and prediction system

Zengchao Hao, Amir Aghakouchak, Navid Nakhjiri & Alireza Farahmand

Data in figshare
Code in figshare
Integrated figshare data viewer

Cited 37 times, according to Google Scholar
Thank you for listening!

andrew.hufton@nature.com

Thank you to colleagues:

Iain Hrynaszkiewicz, Head of Data and HSS Publishing
Sowmya Swaminathan, Head of Editorial Policy