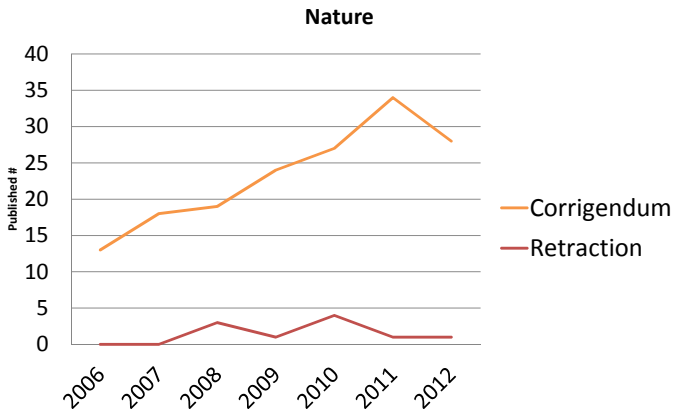
The importance of journals and universities in tackling irreproducibility

Philip Campbell
PCAST Washington DC
31 Jan 2014

Growth in formal corrections



Growth in formal corrections

(Examples from Nature, Nature Biotechnology, Nature Neuroscience, Nature Methods)

- Missing controls, results not sufficiently representative of experimental variability, data selection
- Investigator bias, e.g., in determining the boundaries of an area to study (lack of blinding)
- Technical replicates wrongly described as biological replicates
- Over-fitting of models for noisy datasets in various experimental settings: fMRI, x-ray crystallography, machine learning
- Errors and inappropriate manipulation in image presentation, poor data management
- Contamination of primary culture cells

Mandating reporting standards is not sufficient

MIAME – Minimal Information About a Microarray Experiment

2002: Nature journals mandate deposition of MIAME-compliant microarray data
2006: compliance issues identified

Ioannidis *et al.,* Nat Gen 41, 2, 149 (2009)

Repeatability of published microarray gene expression analyses

John P A Ioannidis[1–3], David B Allison[4], Catherine A Ball[5], Issa Coulibaly[4], Xiangqin Cui[4], Aedín C Culhane[6,7], Mario Falchi[8,9], Cesare Furlanello[10], Laurence Game[11], Giuseppe Jurman[10], Jon Mangion[11], Tapan Mehta[4], Michael Nitzberg[5], Grier P Page[4,12], Enrico Petretto[11,13] & Vera van Noort[14]

Of 18 papers containing microarray data published in NG in 2005-2006, 10 analyses could not be reproduced, 6 only partially.

Irreproducibility: NPG actions so far

- Awareness raising – meetings 2013/14: NINDS, NCI, Academy of Medical Sciences, Royal Society, Science Europe,……
- Awareness raising: Editorials, articles by experts
- We removed length limits on online methods sections
- We substantially increased figure limits in Nature and improved access to Supp Info data in research journals.
- Agreement with Figshare to present data behind figures
- Transparency: we are considering developing the author contributions statement
- Statistical advisor (Terry Hyslop) and referees appointed
- 'Reducing our irreproducibility' Editorial + check lists for authors, editors and referees (23 April 2013)

nature publishing group npg

Raising awareness: our content

- Tackling the widespread and critical impact of batch effects in high-throughput data, Leek *et al*., *NRG*, Oct 2010
- How much can we rely on published data on potential drug targets? Prinz et al., *NRDD*, Sep 2011
- The case for open computer programs, Ince *et al.*, *Nature*, Feb 2012
- Raise standards for preclinical cancer research, Begley & Ellis, *Nature,* Mar 2012
- Must try harder – Editorial, *Nature*, Mar 2012
- Face up to false positives, MacArthur, *Nature*, Jul 2012
- Error prone – Editorial, *Nature*, Jul 2012
- Next-generation sequencing data interpretation: enhancing reproducibility and accessibility, Nekrutenko & Taylor, *NRG*, Sep 2012
- A call for transparent reporting to optimize the predictive value of preclinical research. Landis *et al.*, *Nature*, Oct 2012
- Know when your numbers are significant, Vaux, *Nature*, Dec 2012
- Reuse of public genome-wide gene expression data, Rung & Brazma, *NRG*, Feb 2013
- Reducing our irreproducibility – Editorial, *Nature*, May 2013
- Reproducibility: Six red flags for suspect work, Begley, *Nature*, May 2013
- Reproducibility: The risks of the replication drive, Bissell, *Nature*, Nov 2013

nature publishing group npg

ANNOUNCEMENT

Reducing our irreproducibility

Over the past year, *Nature* has published a string of articles that highlight failures in the reliability and reproducibility of published research (collected and freely available at go.nature.com/huhbyr). The problems arise in laboratories, but journals such as this one compound them when they fail to exert sufficient scrutiny over the results that they publish, and when they do not publish enough information for other researchers to assess results properly.

From next month, *Nature* and the Nature research journals will introduce editorial measures to address the problem by improving the consistency and quality of reporting in life-sciences articles. To ease the interpretation and improve the reliability of published results we will more systematically ensure that key methodological details are reported, and we will give more space to methods sections. We will examine statistics more closely and encourage authors to be transparent, for example by including their raw data.

Central to this initiative is a checklist intended to prompt authors to disclose technical and statistical information in their submissions, and to encourage referees to consider aspects important for research reproducibility (go.nature.com/oloeip). It was developed after discussions with researchers on the problems that lead to irreproducibility, including workshops organized last year by US National Institutes of Health (NIH) institutes. It also draws on published concerns about reporting standards (or the lack of them) and the collective experience of editors at Nature journals.

The checklist is not exhaustive. It focuses on a few experimental and analytical design elements that are crucial for the interpretation of research results but are often reported incompletely. For example, authors will need to describe methodological parameters that can introduce bias or influence robustness, and provide precise characterization of key reagents that may be subject to biological variability, such as cell lines and antibodies. The checklist also consolidates existing policies about data deposition and presentation.

We will also demand more precise descriptions of statistics, and we will commission statisticians as consultants on certain papers, at the editor's discretion and at the referees' suggestion.

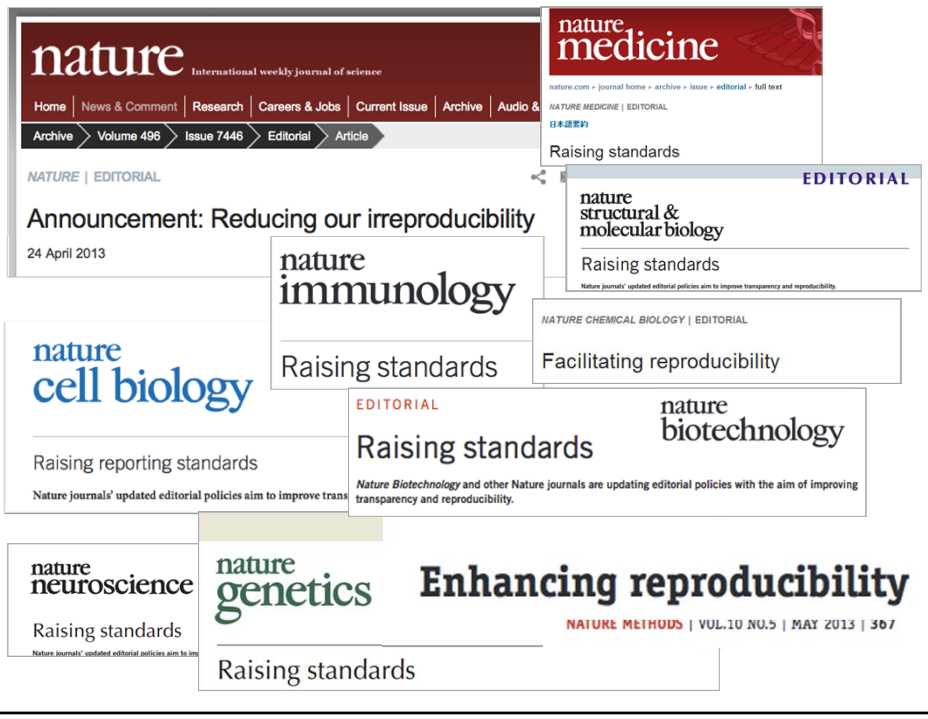
We recognize that there is no single way to conduct an experimental study. Exploratory investigations cannot be done with the same level of statistical rigour as hypothesis-testing studies. Few academic laboratories have the means to perform the level of validation required, for example, to translate a finding from the laboratory to the clinic. However, that should not stand in the way of a full report of how a study was designed, conducted and analysed that will allow reviewers and readers to adequately interpret and build on the results.

To allow authors to describe their experimental design and methods in as much detail as necessary, the participating journals, including *Nature*, will abolish space restrictions on the methods section.

To further increase transparency, we will encourage authors to provide tables of the data behind graphs and figures. This builds on our established data-deposition policy for specific experiments and large data sets. The source data will be made available directly from the figure legend, for easy access. We continue to encourage authors to share detailed methods and reagent descriptions by depositing protocols in Protocol Exchange (www.nature.com/protocolexchange), an open resource linked from the primary paper.

Renewed attention to reporting and transparency is a small step. Much bigger underlying issues contribute to the problem, and are beyond the reach of journals alone. Too few biologists receive adequate training in statistics and other quantitative aspects of their subject. Mentoring of young scientists on matters of rigour and transparency is inconsistent at best. In academia, the ever increasing pressures to publish and chase funds provide little incentive to pursue studies and publish results that contradict or confirm previous papers. Those who document the validity or irreproducibility of a published piece of work seldom get a welcome from journals and funders, even as money and effort are wasted on false assumptions.

Tackling these issues is a long-term endeavour that will require the commitment of funders, institutions, researchers and publishers. It is encouraging that NIH institutes have led community discussions on this topic and are considering their own recommendations. We urge others to take note of these and of our initiatives, and do whatever they can to improve research reproducibility. ■



Reporting Checklist For Life Sciences Articles

This checklist is used to ensure good reporting standards and to improve the reproducibility of published results. For more information, please read Reporting Life Sciences Research.

Figure legends

Each figure legend should contain, for each panel where they are relevant:

- the **exact sample size (*n*)** for each experimental group/condition, given as a number, not a range;
- a **description of the sample collection** allowing the reader to understand whether the samples represent **technical or biological replicates** (including how many animals, litters, cultures, etc.);
- a **statement of how many times the experiment shown was replicated in the laboratory**;
- **definitions of statistical methods and measures**:
  - very common tests, such as *t*-test, simple χ^2 tests, Wilcoxon and Mann-Whitney tests, can be unambiguously identified by name only, but more complex techniques should be described in the methods section;
  - are tests one-sided or two-sided?
  - are there adjustments for multiple comparisons?
  - **statistical test results**, e.g., ***P* values**;
  - definition of '**center values**' as **median or average**;
  - definition of **error bars as s.d. or s.e.m.**

Any descriptions too long for the figure legend should be included in the methods section.

Statistics and general methods

	Reported on page(s) or figure legend(s):
1. How was the sample size chosen to ensure adequate power to detect a pre-specified effect size?	
For animal studies, include a statement about sample size estimate even if no statistical methods were used.	
2. Describe inclusion/exclusion criteria if samples or animals were excluded from the analysis. Were the criteria pre-established?	
3. If a method of randomization was used to determine how samples/animals were allocated to experimental groups and processed, describe it.	
For animal studies, include a statement about randomization even if no randomization was used.	
4. If the investigator was blinded to the group allocation during the experiment and/or when assessing the outcome, state the extent of blinding.	
For animal studies, include a statement about blinding even if no blinding was done.	
5. For every figure, are statistical tests justified as appropriate?	
Do the data meet the assumptions of the tests (e.g., normal distribution)?	
Is there an estimate of variation within each group of data? Is the variance similar between the groups that are being statistically compared?	

Reagents

Reported on page(s) or figure legend(s):

6. To show that antibodies were profiled for use in the system under study (assay and species), provide a citation, catalog number and/or clone number, supplementary information or reference to an antibody validation profile (e.g., Antibodypedia, 1DegreeBio).

7. Identify the source of cell lines and report if they were recently authenticated (e.g., by STR profiling) and tested for mycoplasma contamination.

Animal models

Reported on page(s) or figure legend(s):

8. Report species, strain, sex and age of animals.

9. For experiments involving live vertebrates, include a statement of compliance with ethical regulations and identify the committee(s) approving the experiments.

10. We recommend consulting the ARRIVE guidelines (*PLoS Biol.* **8**(6), e1000412, 2010) to ensure that other relevant aspects of animal studies are adequately reported.

Human subjects

Reported on page(s) or figure legend(s):

11. Identify the committee(s) approving the study protocol.

12. Include a statement confirming that informed consent was obtained from all subjects.

13. For publication of patient photos, include a statement confirming that consent to publish was obtained.

14. Report the clinical trial registration number (at ClinicalTrials.gov or equivalent).

15. For phase II and III randomized controlled trials, please refer to the CONSORT statement and submit the CONSORT checklist with your submission.

16. For tumor marker prognostic studies, we recommend that you follow the REMARK reporting guidelines.

Data deposition

17. Provide accession codes for deposited data.

Reported on page(s) or figure legend(s):

Data deposition in a public repository is mandatory for:
a. Protein, DNA and RNA sequences
b. Macromolecular structures
c. Crystallographic data for small molecules
d. Microarray data

Deposition is strongly recommended for many other datasets for which structured public repositories exist; more details on our data policy are available here. We encourage the provision of other source data in supplementary information or in unstructured repositories such as Figshare and Dryad.

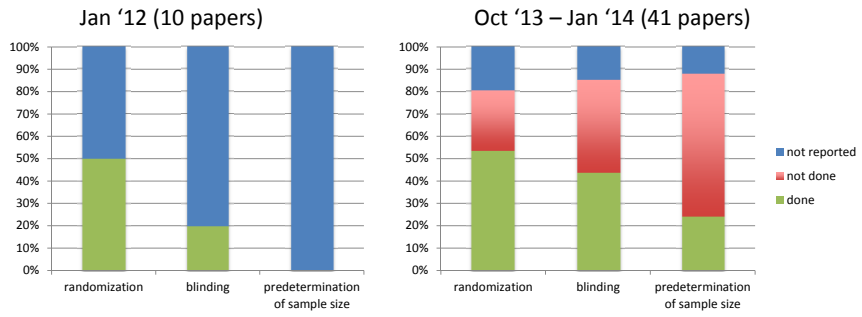
18. Is computer source code provided with the paper or deposited in a public repository? If so, indicate how it can be obtained.

Implementation of reporting checklist

- Onerous!
  - Authors, referees, editors, copyeditors
- Referees:
  - We are not yet sure whether they are paying much attention.
- Authors:
  - Some papers submitted with checklist without prompt
  - Many have embraced source data

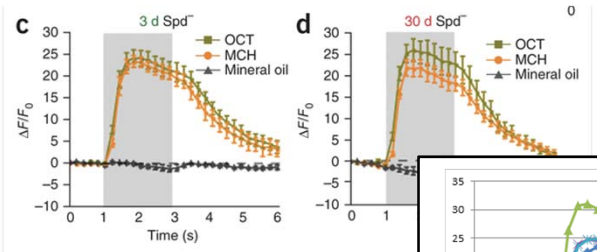
- Improves reporting (see following slides). In May we will conduct a review of the impact, one year on from the checklist's introduction.

Reporting animal experiments in *Nature Neuroscience*



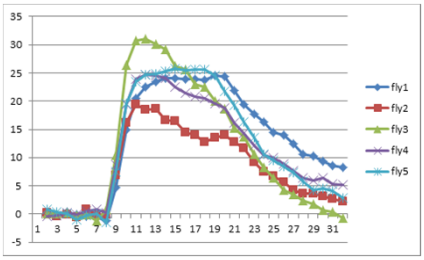
'Not reported' includes cases for which the specific question was not relevant (e.g., investigator cannot be blinded to treatment)
Most frequent problems: power analysis calculations, low n (sample size justification), proper blinding or randomization, multiple t-tests.

Source Data



(**a**) Expression of GCaMP3.0 in the mushroom body of an individual fly, fo
The red line indicates a region of interest used to determine changes in flu
bar represents 50 µm. (**b**) False color–coded image of Ca^{2+} activity in the
lobes shown in **a**. Warm colors indicate high levels, cold colors indicate lo
The numbers indicate changes in fluorescence Δ*F* (%). (**c**,**d**) Time course
mushroom body lobes of 3-d-old and 30-d-old Spd^- flies evoked by the
(MCH) or 3-octanol (OCT) in comparison with the diluent, mineral o
Mann-Whitney *U* test found no substantial difference between
Spd^- flies evoked by the odors). The gray bars indicate the
presented as mean ± s.e.m.

Full size image (281 KB)
Download Excel source data (21 KB)



Allows to plot traces for individual flies from fig 3c with OCT treatment

Previous | Figures/tables index | Next

doi:10.1038/nn.3512 -- Gupta *et al.,* Nat Neuro

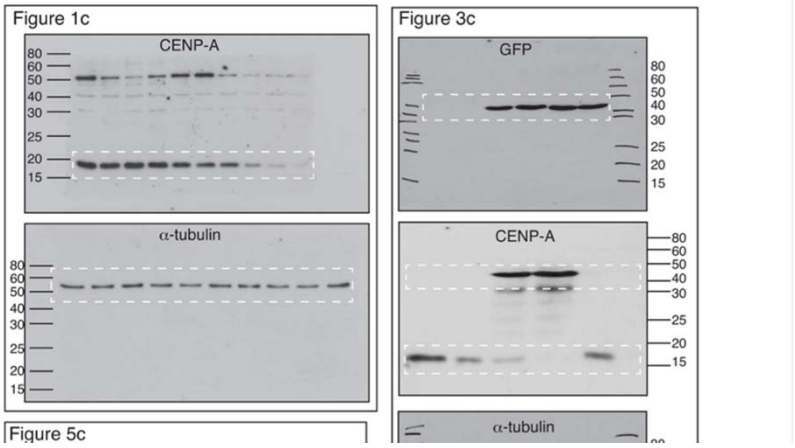
Supplementary Figure 7: Uncropped image of blots from Figs 1c, 3c, 4d, 5c, 6b.

From

A two-step mechanism for epigenetic specification of centromere identity and function

Daniele Fachinetti, H. Diego Folco, Yael Nechemia-Arbely, Luis P. Valente, Kristen Nguyen, Alex J. Wong, Quan Zhu, Andrew J. Holland, Arshad Desai, Lars E. T. Jansen & Don W. Cleveland

Nature Cell Biology **15**, 1056–1066 (2013) | doi:10.1038/ncb2805



Attention needed: Cell line identity

Identify the source of cell lines and indicate if they were recently authenticated (e.g., by STR profiling) and tested for mycoplasma contamination.

This checklist question is not yet enforced as a mandate

Audit of Nature Cell Biology papers (Aug'13 – Dec'13):

- Of 21 relevant papers:
- 20 indicate the source of cell lines(*)
- 4 indicate authentication was done(**)
- 5 acknowledge cell lines were not authenticated
- 17 indicate the cells were tested and demonstrated mycoplasma-free(**)

(*) quality of information variable
(**) timing of tests not always satisfactory

Question about developing author-contribution transparency

- Author contribution statements in Nature journals are informal, unstructured, non-templated.
- Should this change? How? (Possible goals: increased credit, increased accountability for potential flaws.)
- How granular should this information become?

nature publishing group npg

Irreproducibility: underlying issues

- Experimental design: randomization, blinding, sample size determinations, independent experiments vs technical replicates,
- Statistics
- Big data, overfitting (needs gut scepticism/tacit knowledge)
- Gels, microscopy images,
- Reagents validity – antibodies, cell lines
- Animal studies description
- Methods description
- Data deposition

- **Publication bias and refutations – where?**
- **IP confidentiality – replication failures unpublishable**
- **Lab supervision**
- **Lab training**
- **Pressure to publish**
- **"It pays to be sloppy"**

nature publishing group npg

Key challenge: motivating, publishing and highlighting replications and refutations

- Providing a publication channel in original journal and/or elsewhere will take more effort from publishers and editors.
- Formal misconduct investigations will sometimes lead to retractions, but many papers that have been shown to be fundamentally flawed will not be explicitly refuted, even if a channel exists.
- In practical terms, it is challenging for explicit refutations and replications to be identified and highlighted and linked from the original paper. We are looking into this.

nature publishing group npg

Universities/institutes: target issues

- Data validation
- Lab size and management
- Training
- Publication bias
- Data/notebooks access
- Reagent access

nature publishing group npg

Part of a possible training course

- **DAY 1**
- 9h00 – 10h00 Introductions
- 10h00 – 11h00 Case study
- 11h00 – 11h30 Break
- 11h30 – 13h00 Mentoring
- 13h00 – 14h00 Lunch break
- 14h00 – 14h45 Diagnosing issues – Vignette 1: Dealing with processed data
- 14h45 – 15h30 Diagnosis issues – Vignette 2: Adventures in data mismanagement
- 15h30 – 16h00 Break
- 16h00 – 16h30 Data management principles and resources

nature publishing group npg

Thanks for listening

nature publishing group npg