On the Reproducibility of Science: Unique Identification of Research Resources in the Biomedical Literature

Nicole A. Vasilevsky[1], Matthew Brush[1], Holly Paddock[2], Laura Ponting[3], Shreejoy Tripathy[4], Greg LaRocca[4], Melissa A. Haendel[1]

[1]Ontology Development Group, Oregon Health & Science University, Portland, OR, [2]ZFIN, University of Oregon, Eugene, OR, [3]. FlyBase, Department of Genetics, University of Cambridge, Cambridge, UK, [4]Department of Biological Sciences and Center for the Neural Basis of Cognition, Carnegie Mellon University, Pittsburgh, PA, USA

Introduction

Despite the proliferation and easy access to scholarly communications, a problem still exists - there is a significant lack of detailed information about the resources reported in publications, which hinders adequate research reproducibility. In cases such as antibodies and model organisms, this lack of unique reference makes it difficult or impossible to reproduce the experiments. In order to better understand the magnitude of this problem, we designed an experiment to evaluate the "identifiability" of research resources in the biomedical literature.

Methods

5 Resource Types:

- Model organisms
- Antibodies
- Cell lines
- Knockdown reagents
- Constructs

5 Domains:

- Cell Biology
- Developmental Biology
- General Biology
- Immunology
- Neuroscience

Example criteria for identifability:

- ☑ Source reported
- ☑ Identifiable in vendor site
- ☑ Identifiable in MOD
- ☑ Catalog number reported

3 impact factors

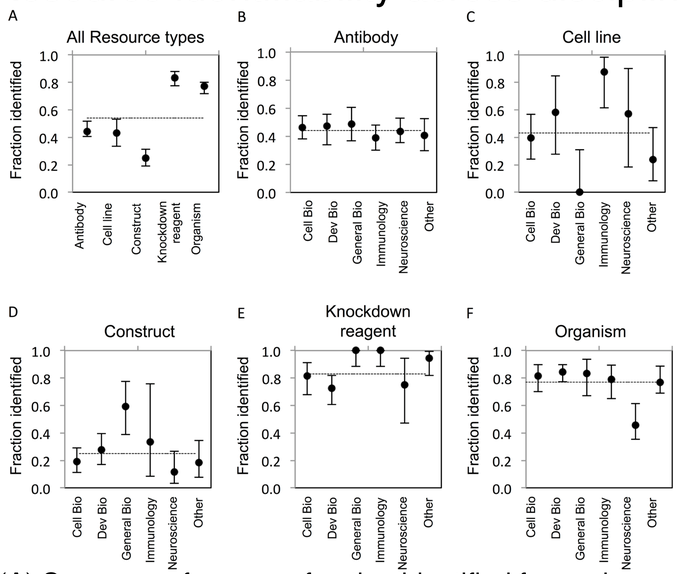
- High impact
- Mid impact
- Low impact

3 Reporting guidelines

- Stringent
- Satisfactory
- Loose

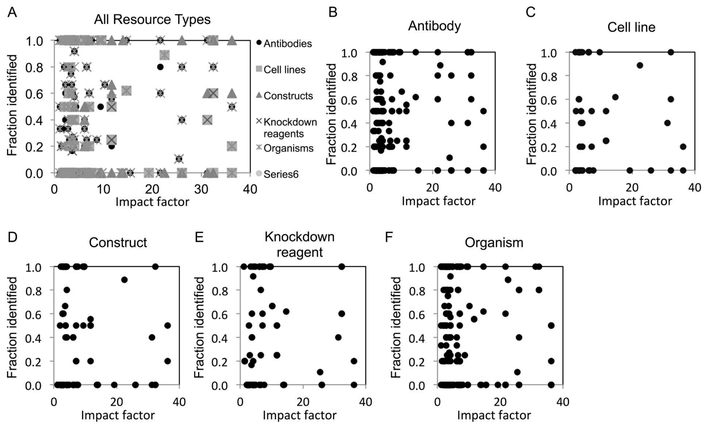
238 papers were curated

Resource identifiability across disciplines



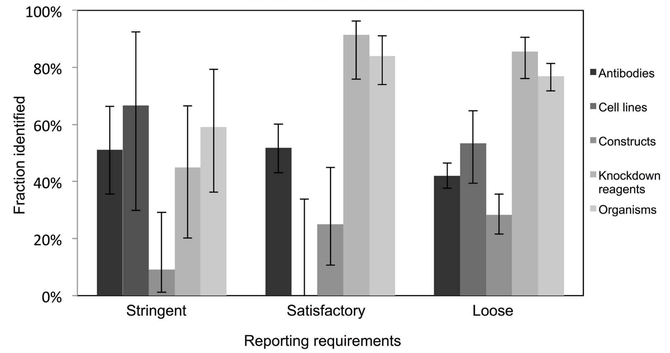
(A) Summary of average fraction identified for each resource type. (B–F) Identifiability of each resource type by discipline.

Resource identification rates across journals of varying impact factors



(A) An overview of fraction identified by impact factor for all resource types. (B–F) Fraction identified by impact factor for each individual resource type. Increasing height on the x-axis corresponds with a higher impact factor for each journal.

Stringent resource reporting requirements does not improve resource identification



The reporting requirements for each journal were classified as stringent, satisfactory or loose. A total of 53 out of 118 resources were identifiable in the stringent reporting guidelines category, 201 resources were identifiable out of 329 resources for the satisfactory category and 662 out of 1,217 resources were identifiable in the loose category.

Recommended reporting guidelines for life science resources

FORCE11 The Future of Research Communications and e-Scholarship
http://www.force11.org/node/4433

biosharing.org
http://biosharing.org/bsg-000532

Conclusions:

- ❑ Inability to identify resources hinders reproducibility
- ✓ Improve metadata standards for tracking resources, authors should provide unique IDs in publications
- ❑ Current reporting standards are insufficient to uniquely identify resources
- ✓ Publishers, editors, and reviewers should work together to increase reporting requirements

Funding: OHSU acknowledges the support of the OHSU Library and #1R24OD011883-01 from the NIH Office of the Director. Holly Paddock and Laura Ponting are funded grant #'s P41 HG002659 and P41 HG000739, respectively. Shreejoy Tripathy is funded by an NSF graduate research fellowship and a RK Mellon Foundation fellowship. Greg LaRocca is funded by NIH grants R01DC005798 and R01DC011184.