The bioRxiv Preprint Server:
An Open Science Initiative for the Life Sciences

John Inglis Ph.D.
Co-founder at bioRxiv
Cold Spring Harbor Laboratory
twitter @JohnRInglis
inglis@cshl.edu

National Academy, Washington DC,
September 18, 2017

Preprint (n):

a complete but unpublished manuscript yet to be certified by peer review, distributed by its author before or at submission to a journal

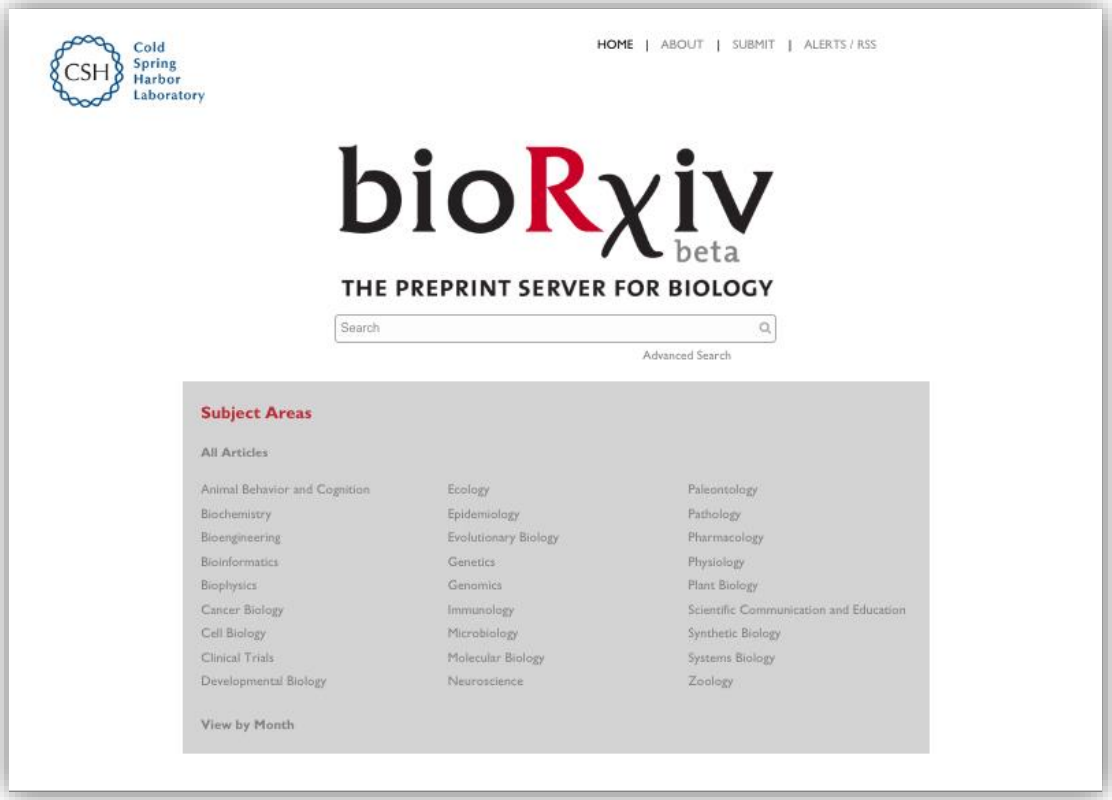
Preprint server (n):

a journal-independent mechanism for distributing preprints in a defined domain

The purpose of preprint distribution

To share the results of recent research freely and openly before they are certified by peer review, in a manner that permits immediate discovery and discussion of the results and feedback to authors from the research community at large

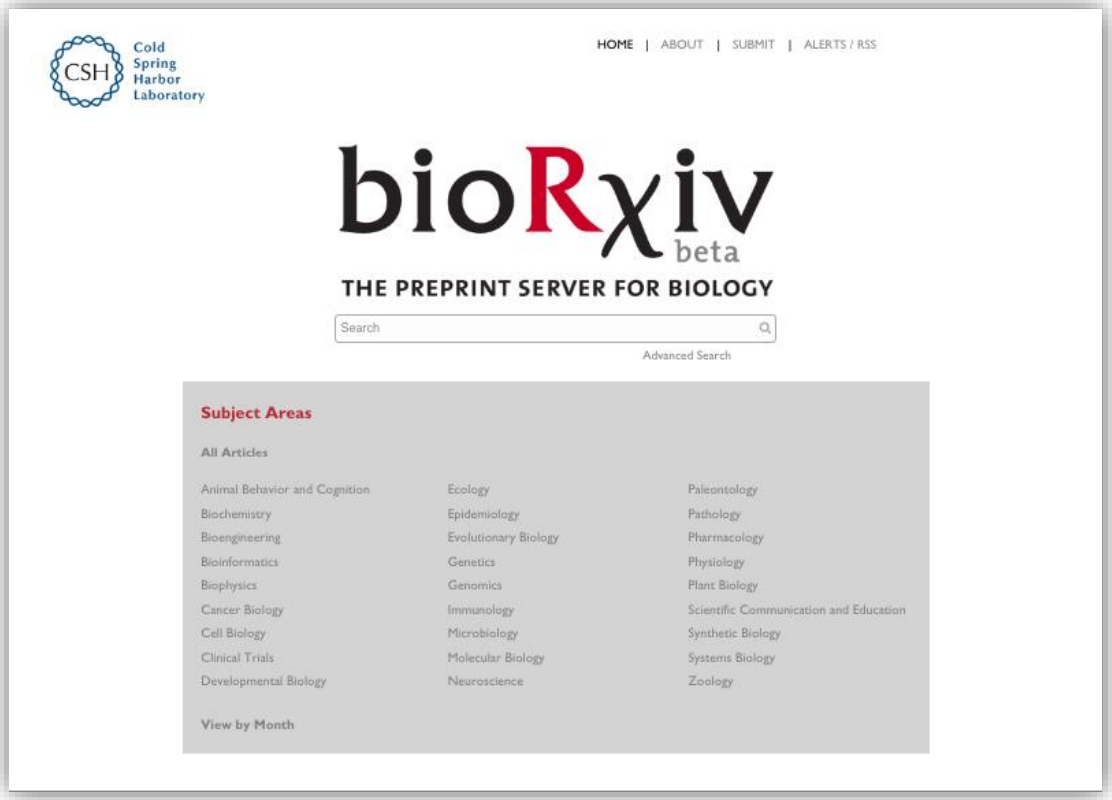
bioRxiv: a server for life science preprints



- Launched November 2013
- Covers all life sciences, clinical trials, epidemiology, and science communication and education
- Operated by Cold Spring Harbor Laboratory
- Hosted by Hi ss
- Supported by

CHAN ZUCKERBERG INITIATIVE

bioRxiv: a server for life science preprints



- Modeled conceptually on arXiv but different tech, features, and functions
- Not-for-profit
- Free for authors and readers
- No reader registration required
- The largest repository of life science preprints

Cold Spring Harbor Laboratory, NY

Posting to bioRxiv: author's choices

1 **Analysis of protein-coding genetic variation in 60,706 humans**

2 Exome Aggregation Consortium[#], Monkol Lek[1,2,3,4], Konrad J Karczewski[1,2*], Eric V
3 Minikel[1,2,5*], Kaitlin E Samocha[1,2,6,5*], Eric Banks[2], Timothy Fennell[2], Anne H O'Donnell-
4 Luria[1,2,7], James S Ware[2,8,9,10,11], Andrew J Hill[1,2,12], Beryl B Cummings[1,2,5], Taru
5 Tukiainen[1,2], Daniel P Birnbaum[2], Jack A Kosmicki[1,2,6,13], Laramie E Duncan[1,2,6], Karol

New, confirmatory, or contradictory results?

A subject category?

A CC license (Zero, BY, BY-NC, BY-ND, BY-NC-ND) or reserve all rights?

Posting to bioRxiv: screening

1 **Analysis of protein-coding genetic variation in 60,706 humans**

2 Exome Aggregation Consortium[#], Monkol Lek[1,2,3,4], Konrad J Karczewski[1,2*], Eric V
3 Minikel[1,2,5*], Kaitlin E Samocha[1,2,6,5*], Eric Banks[2], Timothy Fennell[2], Anne H O'Donnell-
4 Luria[1,2,7], James S Ware[2,8,9,10,11], Andrew J Hill[1,2,12], Beryl B Cummings[1,2,5], Taru
5 Tukiainen[1,2], Daniel P Birnbaum[2], Jack A Kosmicki[1,2,6,13], Laramie E Duncan[1,2,6], Karol

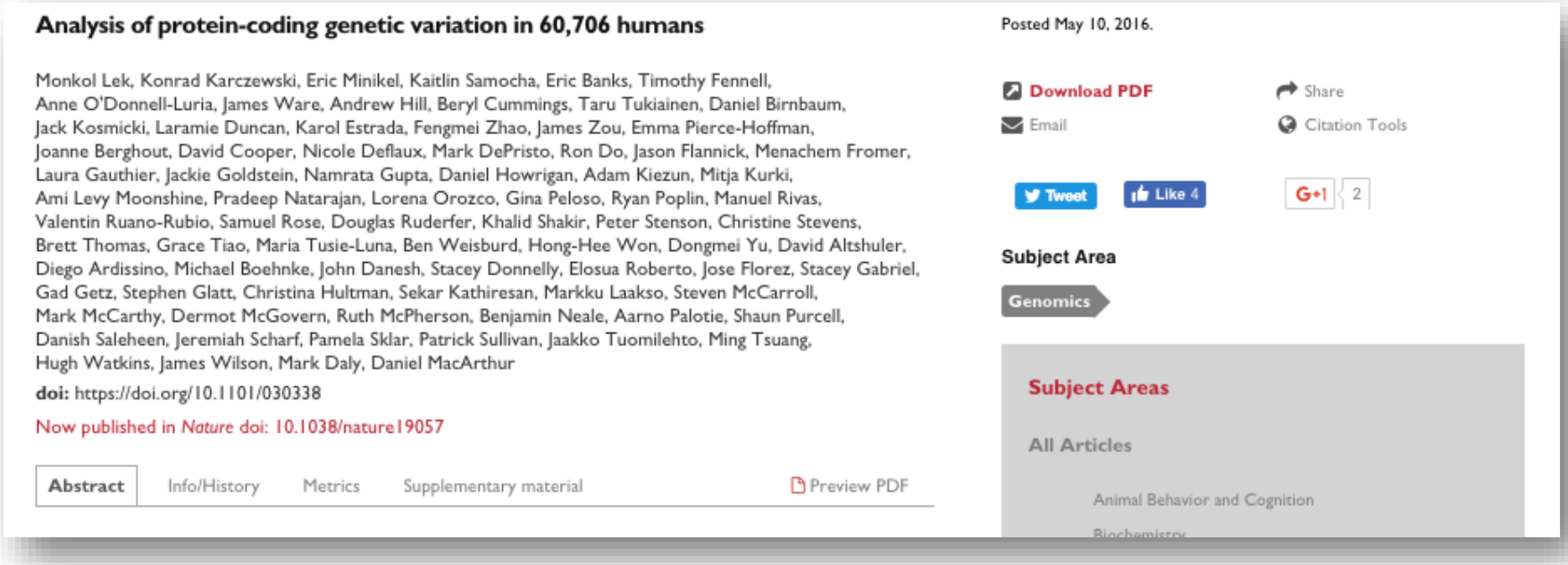
1. **bioRxiv staff check for**:

- Appropriate scope
- Plagiarism
- Images of human subjects

2. **Scientific and clinical affiliates check for**:

- Non-science
- Appropriate intent
- Human health implications

Posting to bioRxiv: display



Analysis of protein-coding genetic variation in 60,706 humans

Monkol Lek, Konrad Karczewski, Eric Minikel, Kaitlin Samocha, Eric Banks, Timothy Fennell, Anne O'Donnell-Luria, James Ware, Andrew Hill, Beryl Cummings, Taru Tukiainen, Daniel Birnbaum, Jack Kosmicki, Laramie Duncan, Karol Estrada, Fengmei Zhao, James Zou, Emma Pierce-Hoffman, Joanne Berghout, David Cooper, Nicole Deflaux, Mark DePristo, Ron Do, Jason Flannick, Menachem Fromer, Laura Gauthier, Jackie Goldstein, Namrata Gupta, Daniel Howrigan, Adam Kiezun, Mitja Kurki, Ami Levy Moonshine, Pradeep Natarajan, Lorena Orozco, Gina Peloso, Ryan Poplin, Manuel Rivas, Valentin Ruano-Rubio, Samuel Rose, Douglas Ruderfer, Khalid Shakir, Peter Stenson, Christine Stevens, Brett Thomas, Grace Tiao, Maria Tusie-Luna, Ben Weisburd, Hong-Hee Won, Dongmei Yu, David Altshuler, Diego Ardissino, Michael Boehnke, John Danesh, Stacey Donnelly, Elosua Roberto, Jose Florez, Stacey Gabriel, Gad Getz, Stephen Glatt, Christina Hultman, Sekar Kathiresan, Markku Laakso, Steven McCarroll, Mark McCarthy, Dermot McGovern, Ruth McPherson, Benjamin Neale, Aarno Palotie, Shaun Purcell, Danish Saleheen, Jeremiah Scharf, Pamela Sklar, Patrick Sullivan, Jaakko Tuomilehto, Ming Tsuang, Hugh Watkins, James Wilson, Mark Daly, Daniel MacArthur

doi: https://doi.org/10.1101/030338

Now published in *Nature* doi: 10.1038/nature19057

Abstract | Info/History | Metrics | Supplementary material | Preview PDF

Posted May 10, 2016.

Download PDF | Share | Email | Citation Tools

Tweet | Like 4 | G+1 2

Subject Area

Genomics

Subject Areas

All Articles

Animal Behavior and Cognition

Biochemistry

Manuscript goes live within 24 hours of posting

DOI and unique URL assigned

Downloadable PDF

Updatable at any time: all versions remain available

Posting to bioRxiv: display

Analysis of protein-coding genetic variation in 60,706 humans

Monkol Lek, Konrad Karczewski, Eric Minikel, Kaitlin Samocha, Eric Banks, Timothy Fennell, Anne O'Donnell-Luria, James Ware, Andrew Hill, Beryl Cummings, Taru Tukiainen, Daniel Birnbaum, Jack Kosmicki, Laramie Duncan, Karol Estrada, Fengmei Zhao, James Zou, Emma Pierce-Hoffman, Joanne Berghout, David Cooper, Nicole Deflaux, Mark DePristo, Ron Do, Jason Flannick, Menachem Fromer, Laura Gauthier, Jackie Goldstein, Namrata Gupta, Daniel Howrigan, Adam Kiezun, Mitja Kurki, Ami Levy Moonshine, Pradeep Natarajan, Lorena Orozco, Gina Peloso, Ryan Poplin, Manuel Rivas, Valentin Ruano-Rubio, Samuel Rose, Douglas Ruderfer, Khalid Shakir, Peter Stenson, Christine Stevens, Brett Thomas, Grace Tiao, Maria Tusie-Luna, Ben Weisburd, Hong-Hee Won, Dongmei Yu, David Altshuler, Diego Ardissino, Michael Boehnke, John Danesh, Stacey Donnelly, Elosua Roberto, Jose Florez, Stacey Gabriel, Gad Getz, Stephen Glatt, Christina Hultman, Sekar Kathiresan, Markku Laakso, Steven McCarroll, Mark McCarthy, Dermot McGovern, Ruth McPherson, Benjamin Neale, Aarno Palotie, Shaun Purcell, Danish Saleheen, Jeremiah Scharf, Pamela Sklar, Patrick Sullivan, Jaakko Tuomilehto, Ming Tsuang, Hugh Watkins, James Wilson, Mark Daly, Daniel MacArthur

doi: https://doi.org/10.1101/030338

Now published in *Nature* doi: 10.1038/nature19057

Abstract | Info/History | Metrics | Supplementary material | Preview PDF

Posted May 10, 2016.

Download PDF | Share | Email | Citation Tools

Tweet | Like 4 | G+1 2

Subject Area

Genomics

Metrics of usage

Supplementary material/data

Link is made to published version

Posts to bioRxiv

15,000 manuscripts

28% revised

83,000 authors

5400 institutions

101 countries

- 44% USA
- 12% UK
- 6% Germany
- 4% Canada
- 4% France
- 3% Australia
- 2% China

Every subject category

© Science AAS

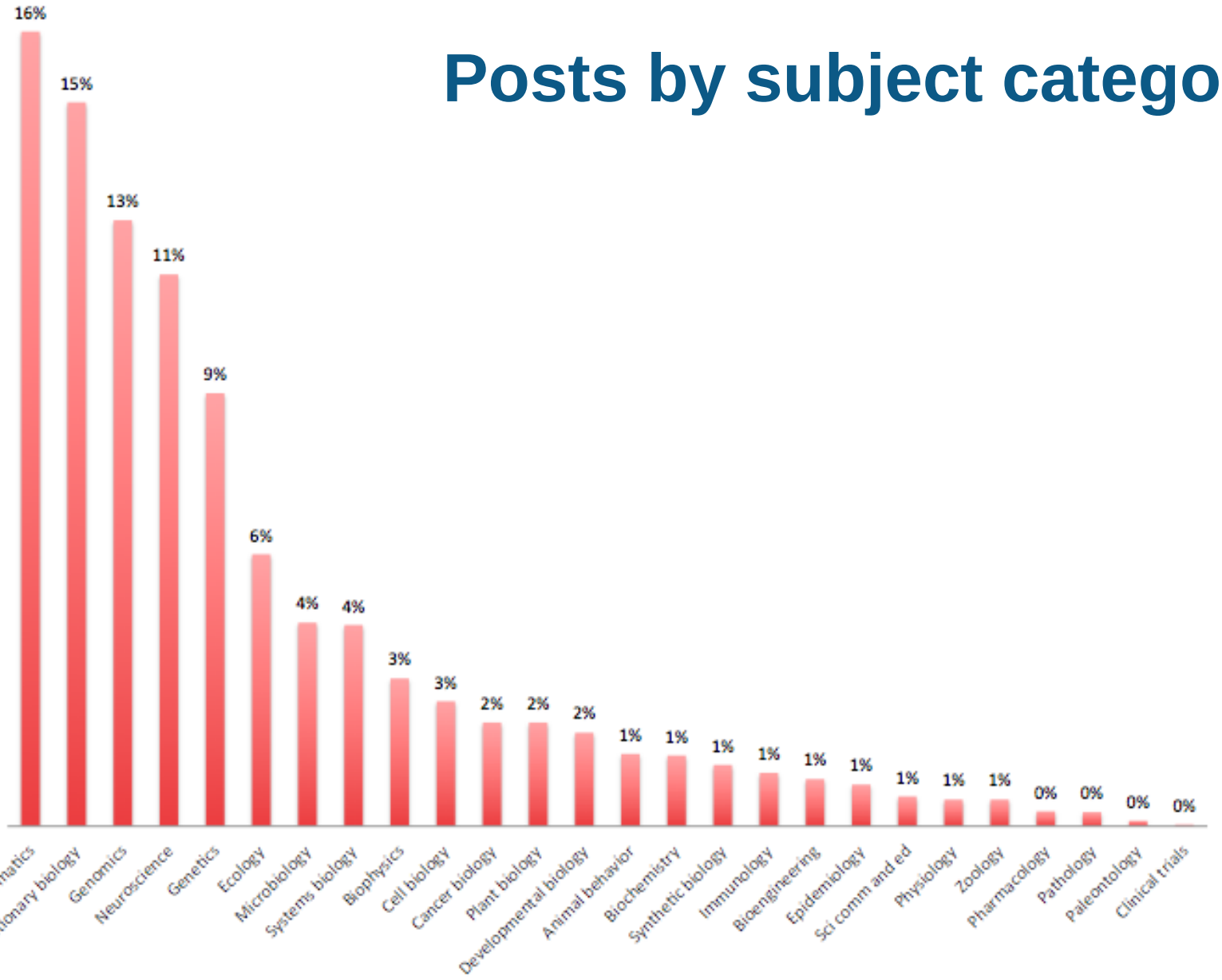
95% of submitted manuscripts accepted

95% “new”, 3% “confirmatory”, 2% “contradictory”

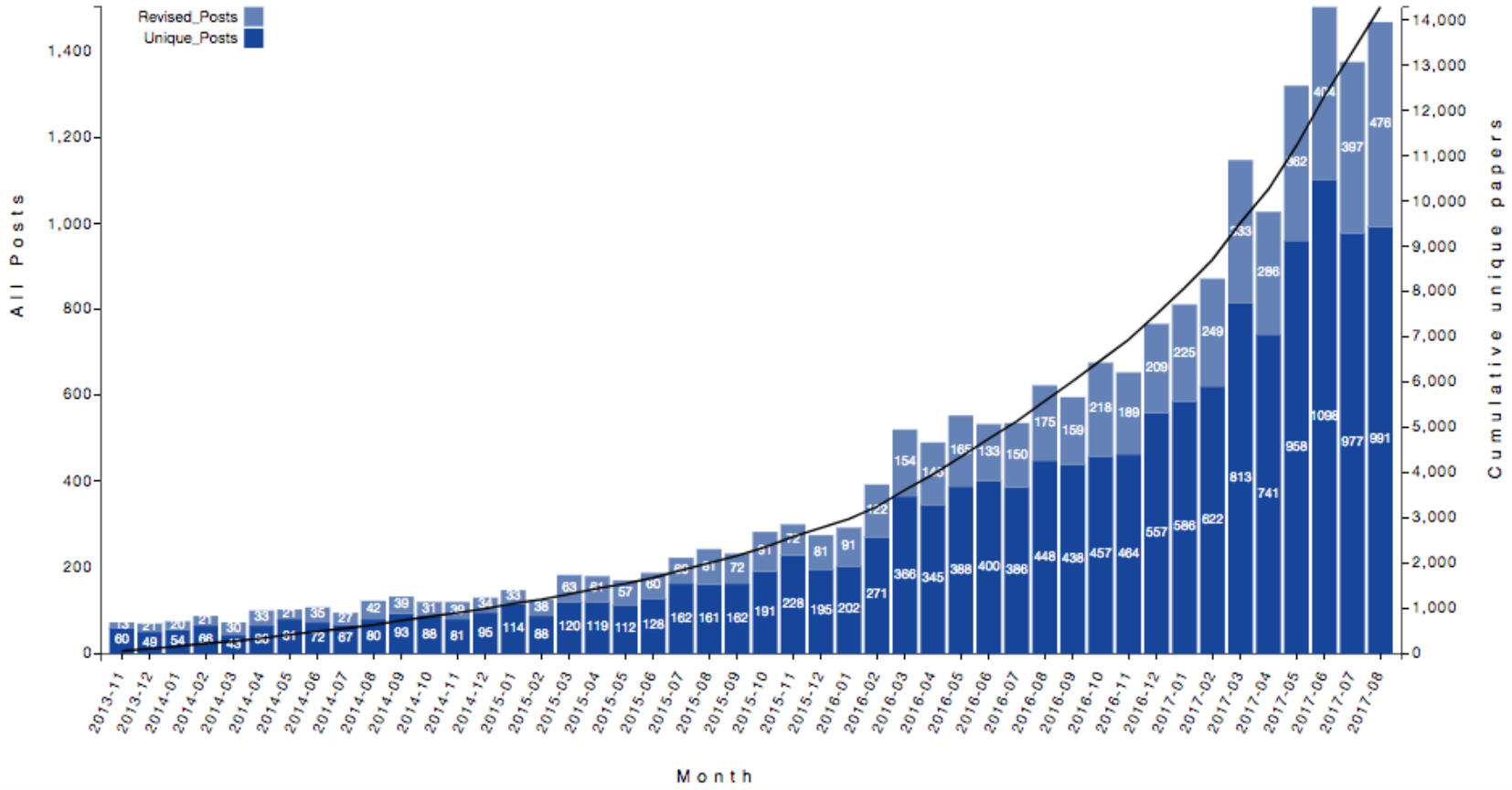
Posts by institution

171	Stanford University
133	University of Cambridge
123	University of Oxford
115	University of Washington
107	Harvard University
95	University of Edinburgh
91	University College London
77	Cornell University
74	University of Michigan
73	University of Pennsylvania
68	Imperial College London
63	Harvard Medical School
63	University of Chicago
63	Columbia University
60	Wellcome Trust Sanger Institute
56	Johns Hopkins University
55	University of California, Berkeley
53	Princeton University
53	Cold Spring Harbor Laboratory
53	Massachusetts Institute of Technology

Posts by subject category



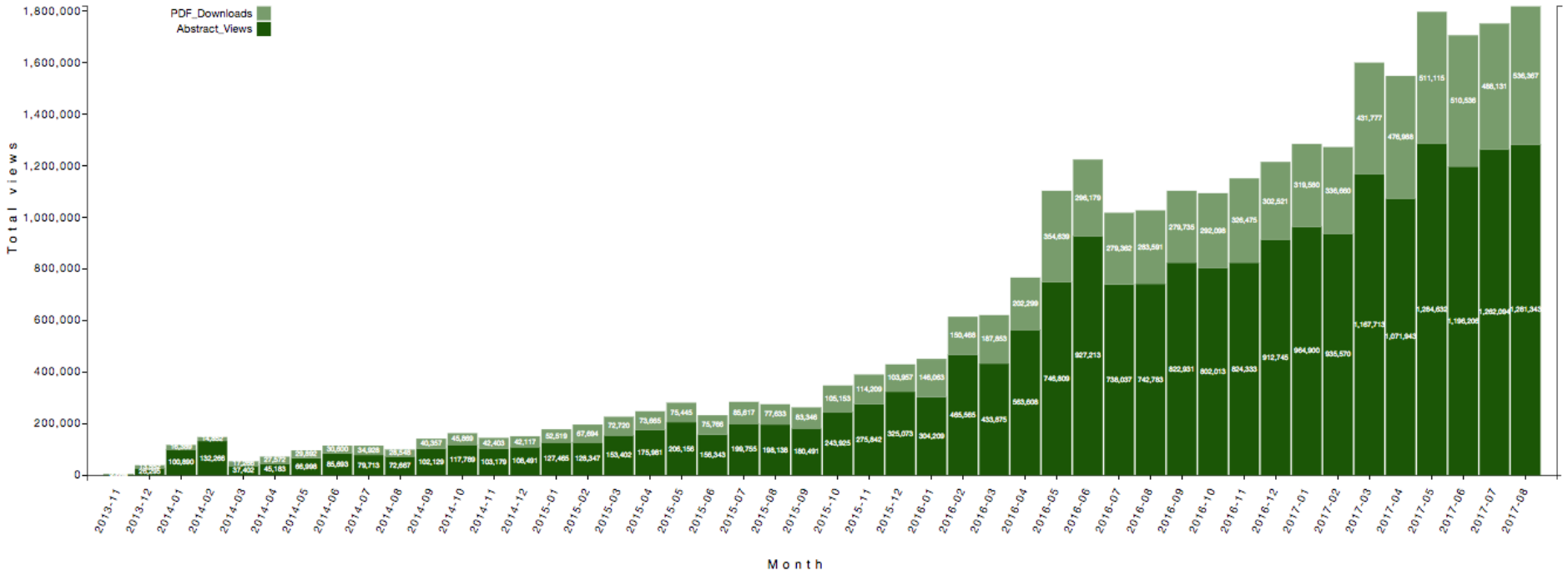
Posts



first submissions: 1098 in June, 977 in July, 991 in August

revised submissions: 404 in June, 379 in July, 476 in August

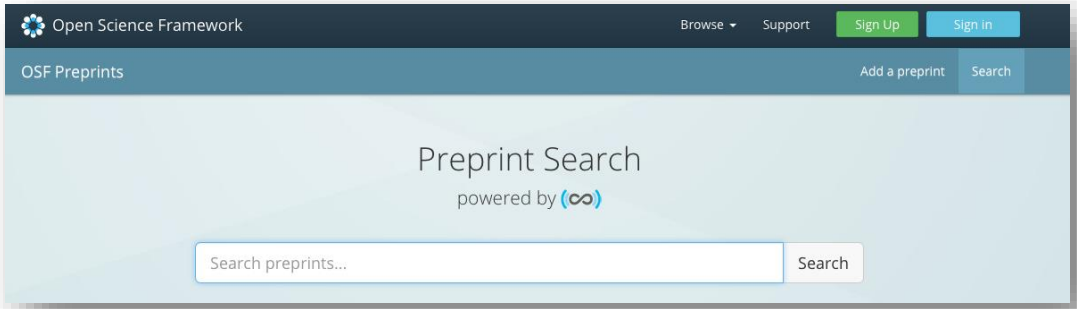
Usage

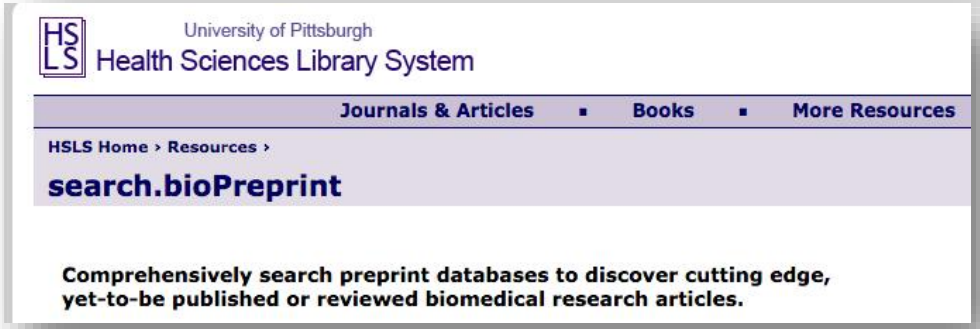


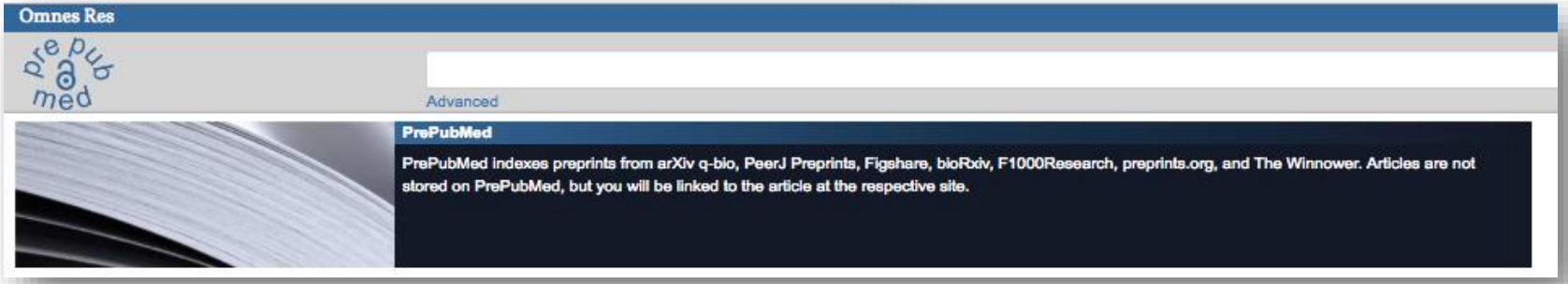
PDF downloads: more than 0.5 million per month

Abstract views: more than 2.1 million per month

Discovery by search

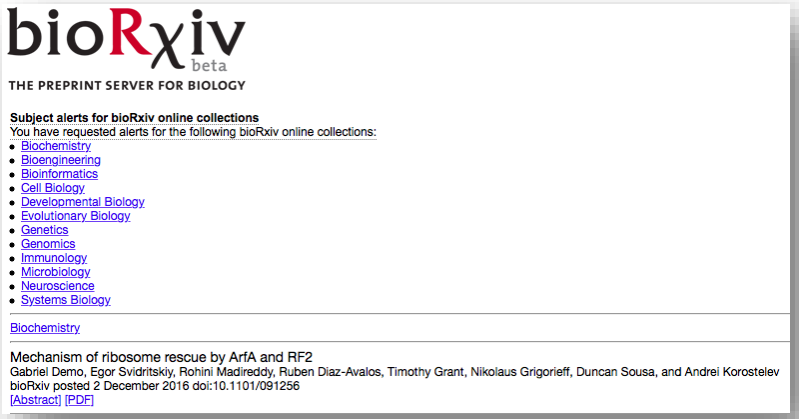






Discovery by alerts

RSS feeds and email alerts
by subject category



Subject category-specific Twitter feeds

Artificial intelligence tools

Preprint integration with journal peer review

Author chooses to submit paper

Journal offers manuscript transfer on submission

2000 transfers to date

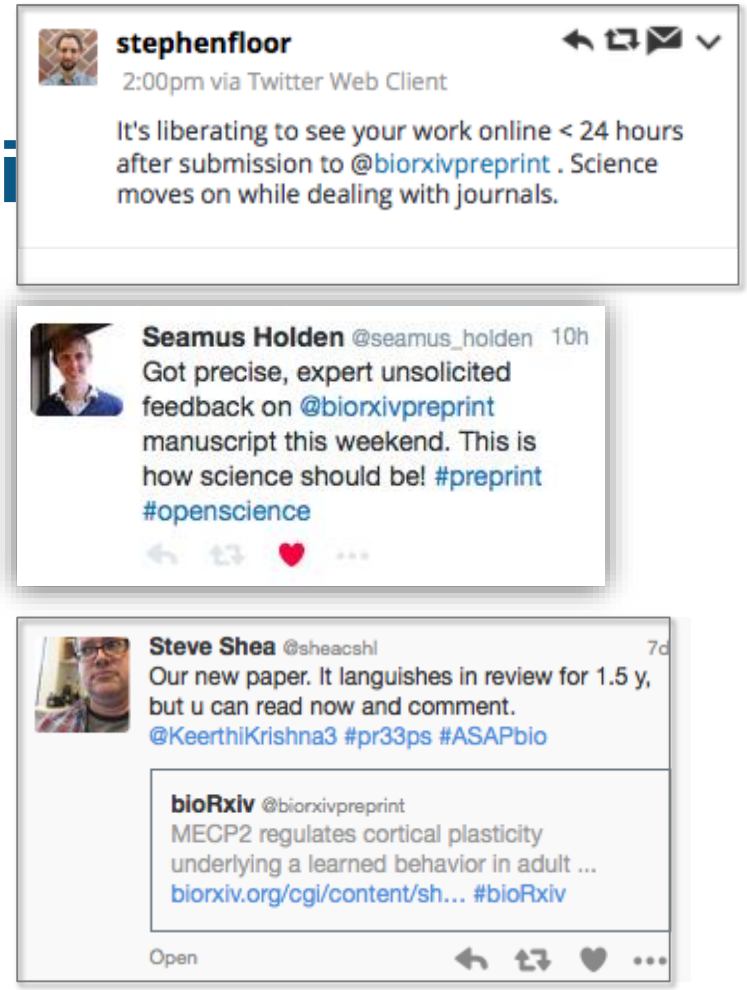
American Association for Immunology
American Society for Biochemistry…
American Society for Cell Biology
American Society for Microbiology
American Society for Plant Biology
American Society of Pharmacology
Biophysical Society
Company of Biologists
Dove Medical Press
eLife
Elsevier
EMBO Press
Genetics Society of America
International Union of Crystallographers
Oxford University Press
Public Library of Science
Rockefeller University Press
Society for Neuroscience
Springer Nature
Wiley Blackwell

Preprint integration with journals

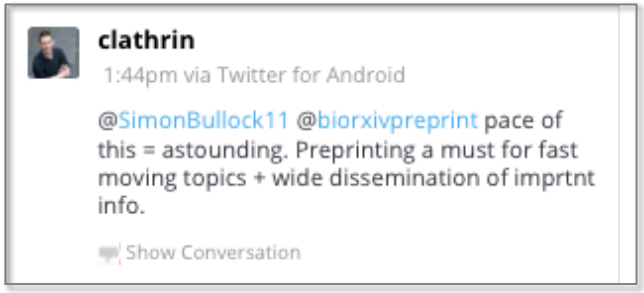
- 2 years after posting, 60% of posted manuscripts have been published in a journal
- 700 journals have published manuscripts first posted to bioRxiv
- These journals vary widely in prestige, specialty, impact factor, commercial and not-for-profit, OA and subscription model
- Many biomedical journals have clarified or changed their attitude to preprin and disapproval is now rare

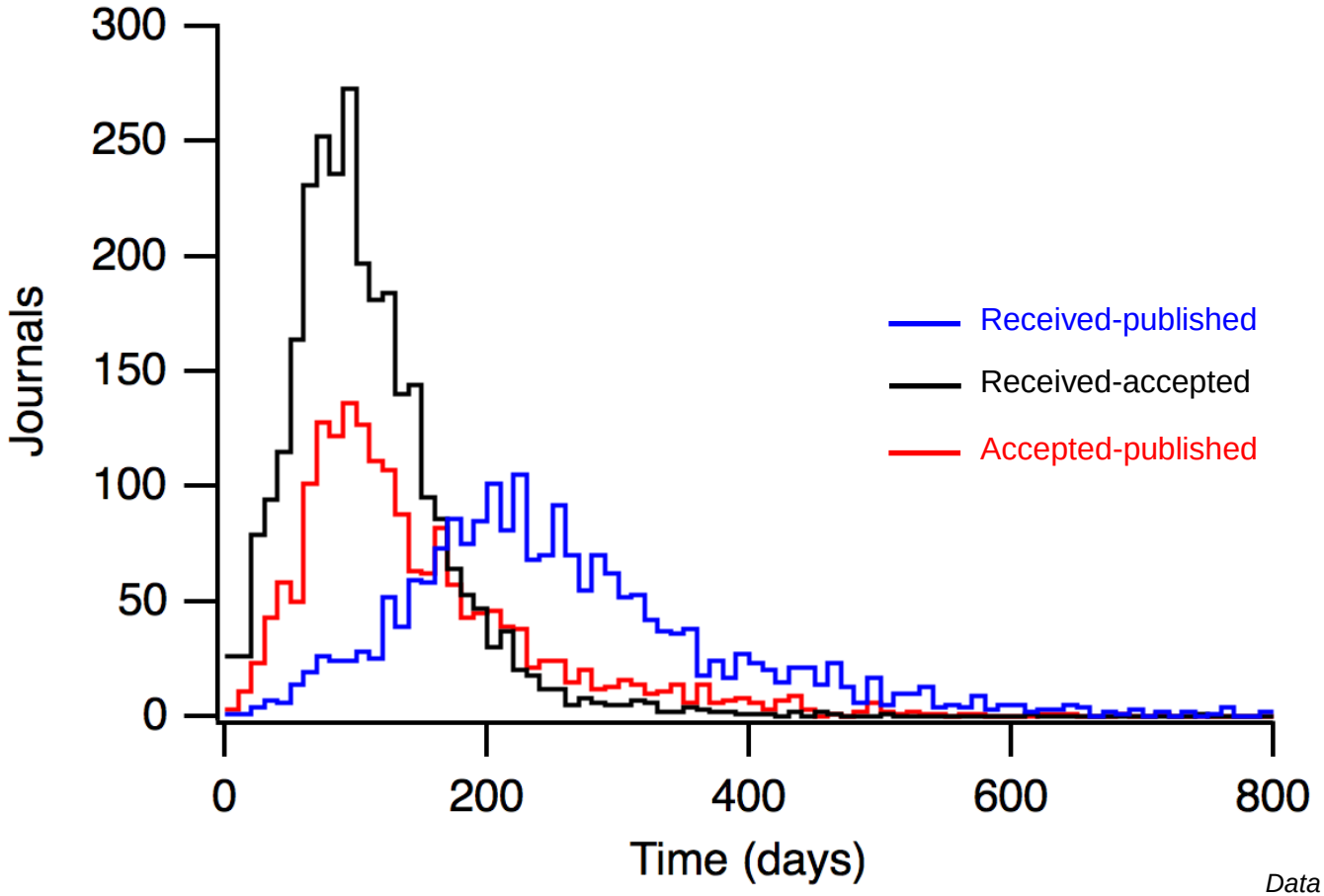
Benefits of bioRxiv prepri

- Rapid and open transmission of results
- Pre-publication feedback/discussion
- Evidence of productivity for grant/hiring committees, especially for early-career investigators
- Dissemination of work that's hard to publish (contradictory or confirmatory) or the author chooses not to publish



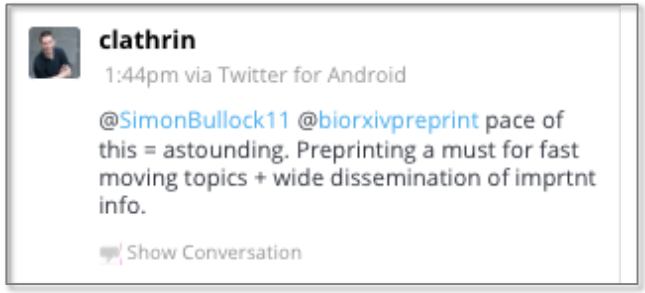
Accelerating communication

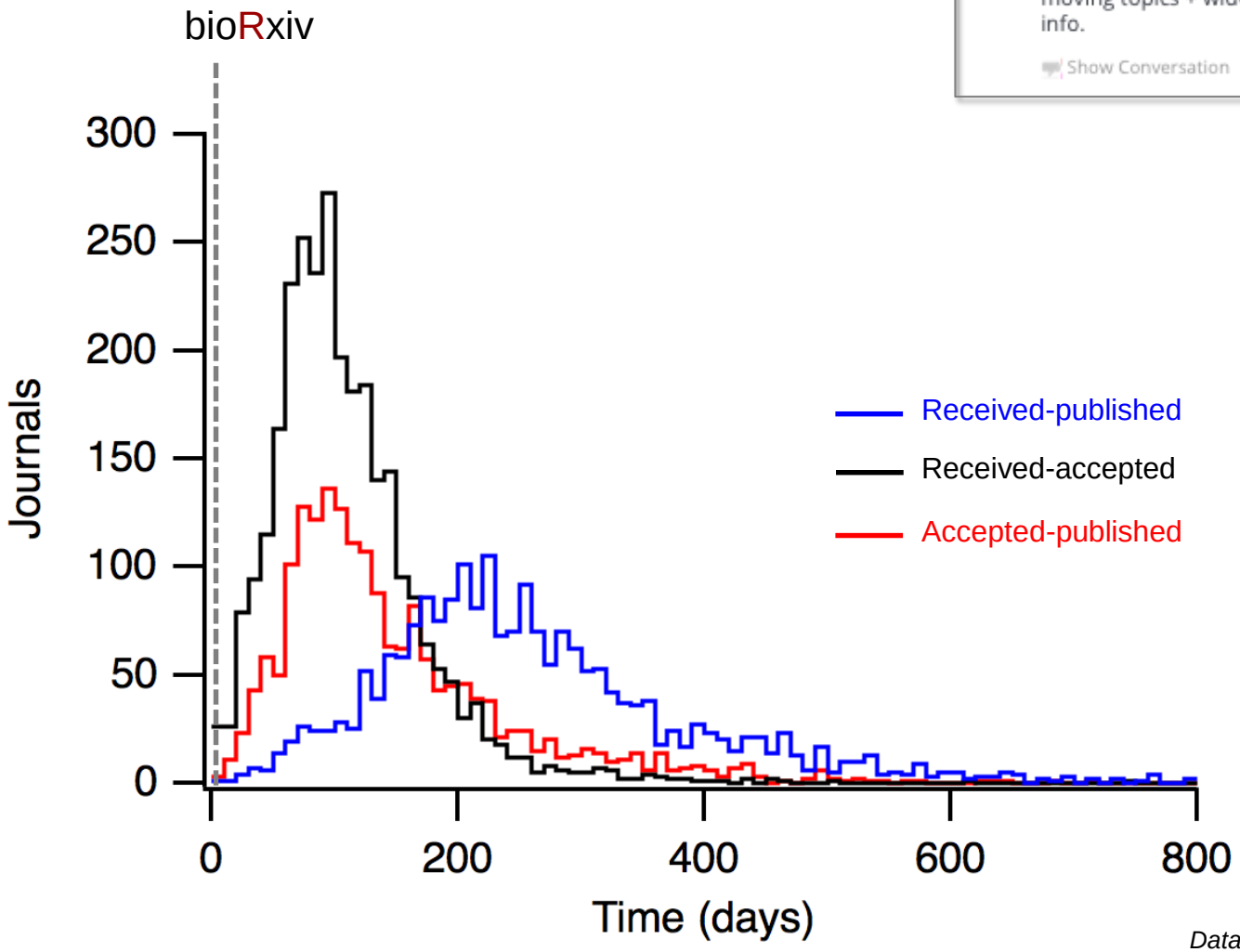




Data courtesy of Stephen Royle

Accelerating communication

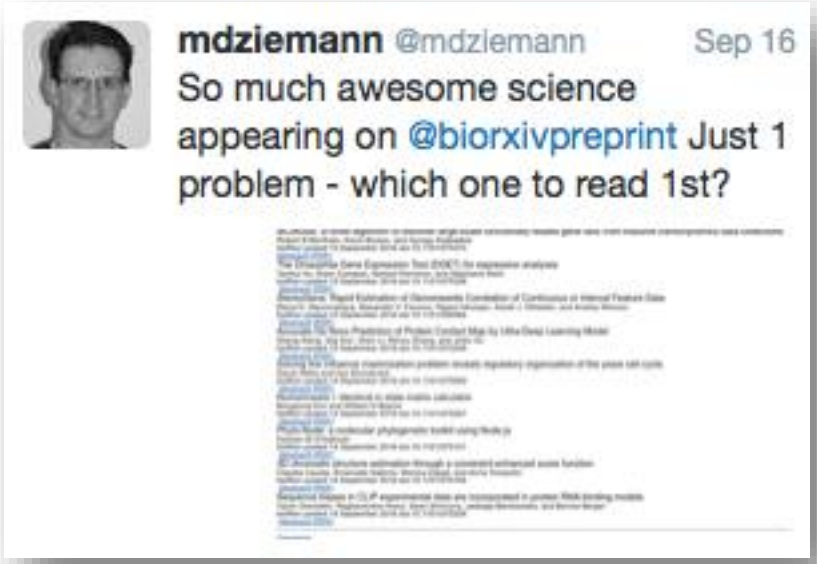




Data courtesy of Stephen Royle

Preprint progress: authors' experience

- They post important science

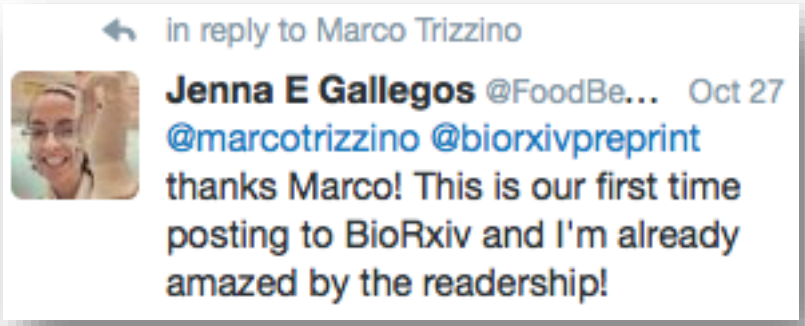


MOST PROLIFIC INSTITUTIONS

- Stanford University
- University of Cambridge
- Harvard University
- University of Oxford
- University of Washington
- University of Edinburgh
- University College London
- Cornell University
- Imperial College London
- University of Chicago

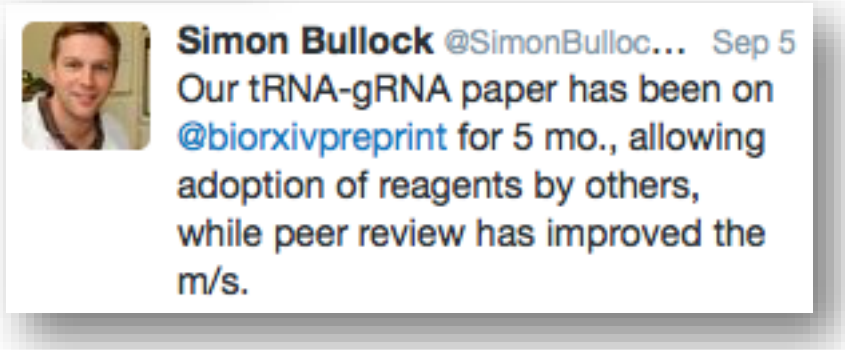
Preprint progress: authors' experience

Their preprints are read



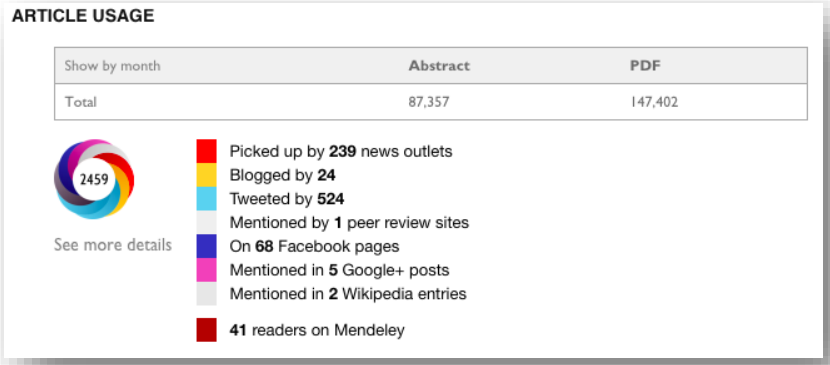
Preprint progress: authors' experience

They get valuable feedback



Preprint progress: impact

Metrics



Blogs and journal clubs

Tweets

twitter

100,000
this year alone

Citations

Cell

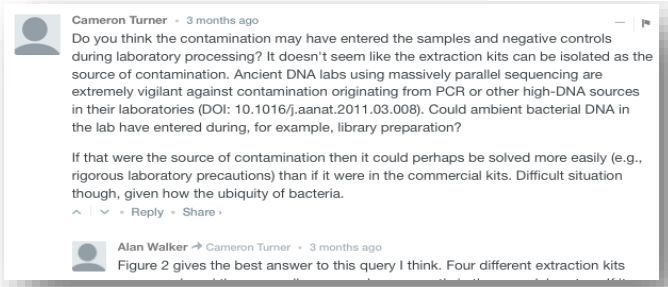
Mangul, S., Olde Loohuis, L.M., Ori, A., Jospin, G., Koslicki, D., Yang, H.T., Wu, T., Boks, M.P., Lomen-Hoerth, C., Wiedau-Pazos, M., et al. (2016). Total RNA Sequencing reveals microbial communities in human blood and disease specific effects. bioRxiv, http://dx.doi.org/10.1101/057570.

Science

H. Ongen, E.T. Dermitzakis, http://biorxiv.org/content/early/2015/01/22/014126 (2015).

Comments

DISQUS



10% of manuscripts

Email/personal contact

Preprint progress: policy changes

- NIH is revising its policies on preprint citation in grant applications

Request for information (RFI): Including Preprints and Interim Research Products in NIH Applications and Reports

grants.nih.gov October 6, 2016

Notice Number: NOT-OD-17-006

- Institutions and funders are accepting preprints as evidence of productivity

Leslie Vosshall @pollyp1 Sep 20

Just back from a review committee meeting where 2 candidates submitted @biorxivpreprint in their materials. And we liked them #ASAPbio

SFARI **SFARI.org**
@SFARIorg

SFARI supports preprints for the life sciences bit.ly/1NCrmZg bit.ly/1rZrAQa #ASAPBio

3:03pm

Use of preprint servers

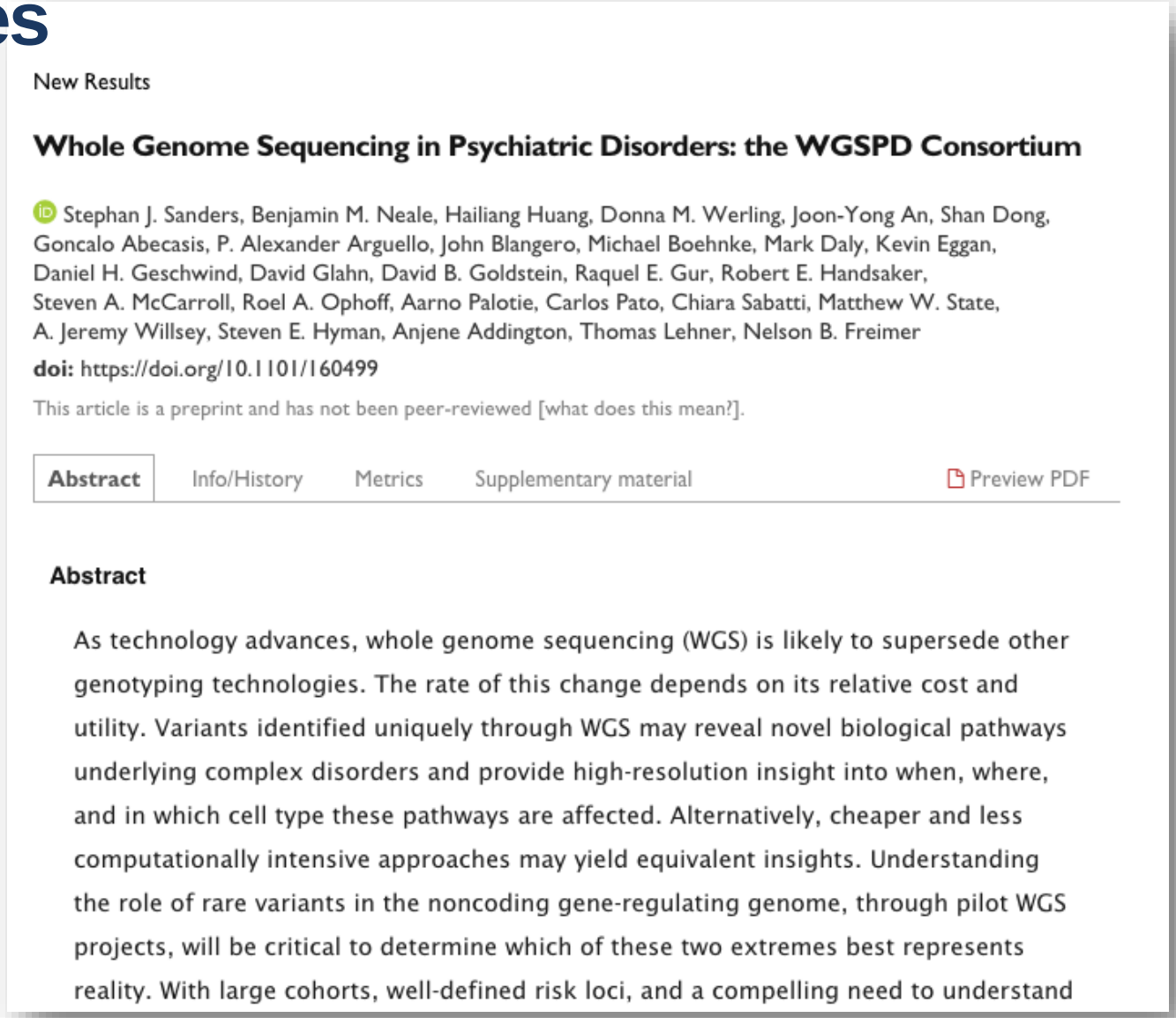
The Board of Trustees of the International Human Frontier Science Program Organization (HFSPO) has decided that for competitions starting in calendar year 2017, applicants may list preprint articles in the publication section of HFSP proposals. Current HFSP awardees are also permitted to cite publications which are deposited in freely available preprint repositories in interim and final reports to the Organization.

New features

Full text XML

Improved submission interface

A content repository for text and data mining



New Results

Whole Genome Sequencing in Psychiatric Disorders: the WGSPD Consortium

Stephan J. Sanders, Benjamin M. Neale, Hailiang Huang, Donna M. Werling, Joon-Yong An, Shan Dong, Goncalo Abecasis, P. Alexander Arguello, John Blangero, Michael Boehnke, Mark Daly, Kevin Eggan, Daniel H. Geschwind, David Glahn, David B. Goldstein, Raquel E. Gur, Robert E. Handsaker, Steven A. McCarroll, Roel A. Ophoff, Aarno Palotie, Carlos Pato, Chiara Sabatti, Matthew W. State, A. Jeremy Willsey, Steven E. Hyman, Anjene Addington, Thomas Lehner, Nelson B. Freimer

doi: https://doi.org/10.1101/160499

This article is a preprint and has not been peer-reviewed [what does this mean?].

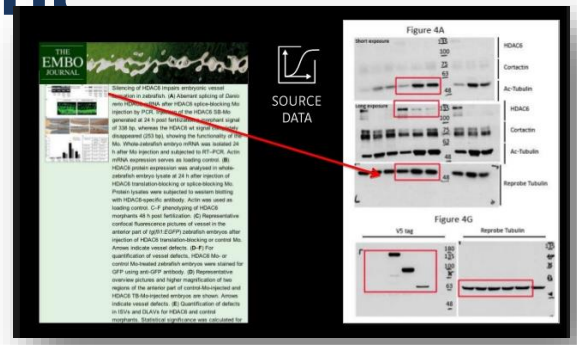
Abstract | Info/History | Metrics | Supplementary material | Preview PDF

Abstract

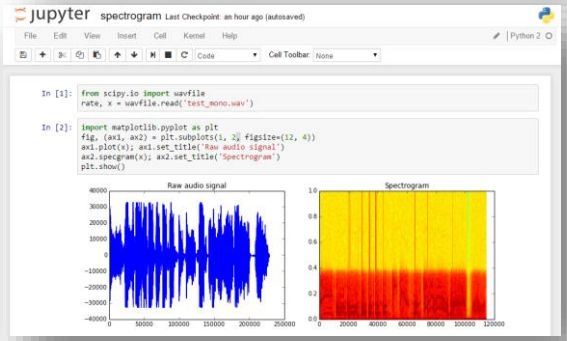
As technology advances, whole genome sequencing (WGS) is likely to supersede other genotyping technologies. The rate of this change depends on its relative cost and utility. Variants identified uniquely through WGS may reveal novel biological pathways underlying complex disorders and provide high-resolution insight into when, where, and in which cell type these pathways are affected. Alternatively, cheaper and less computationally intensive approaches may yield equivalent insights. Understanding the role of rare variants in the noncoding gene-regulating genome, through pilot WGS projects, will be critical to determine which of these two extremes best represents reality. With large cohorts, well-defined risk loci, and a compelling need to understand

Content enrichment

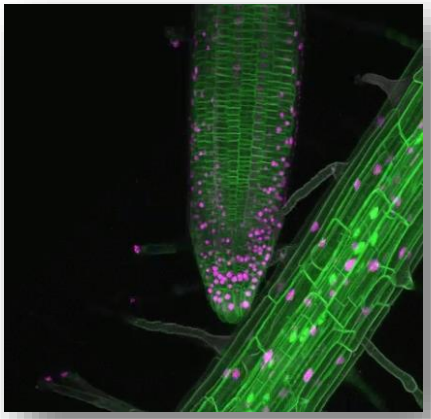
Source data



Jupyter notebooks



Play in place video



Preprints are gaining momentum

- bioRxiv is growing
- arXiv remains unaltered
- chemRxiv has launched this year
- An earth and space sciences server is coming
- A medical arXiv is in development
- There are numerous smaller servers eg in psychology, social science, engineering, sports science,
- There are emerging regional servers eg Indonesia, South America
- Experience reported by authors and readers is overwhelmingly positive

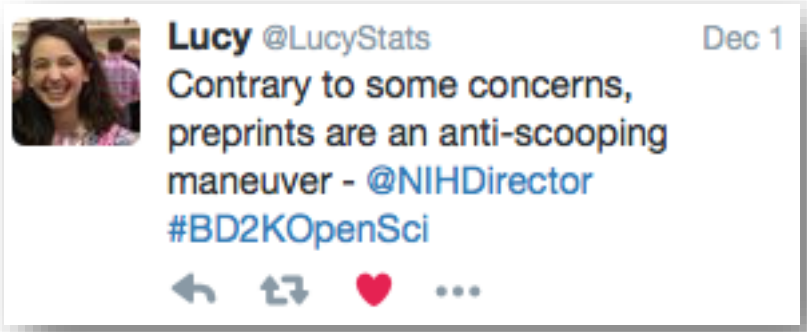
How preprints assist open science

- They are free to post and to read
- They speed up communication
- Scholars can evaluate new findings and their reliability
  without the delay introduced by journal peer review
- Scholars can now post preprints without anxiety about publication
- Funders are providing incentives to those who preprint
- Contradictory and confirmatory results can be posted
- Text and data mining is permitted by bioRxiv terms of use

How preprints assist open science: obstacles

- Anxieties about scooping i.e being first to publish in a journal

- Yet preprints can prevent scooping



How preprints assist open science: obstacles

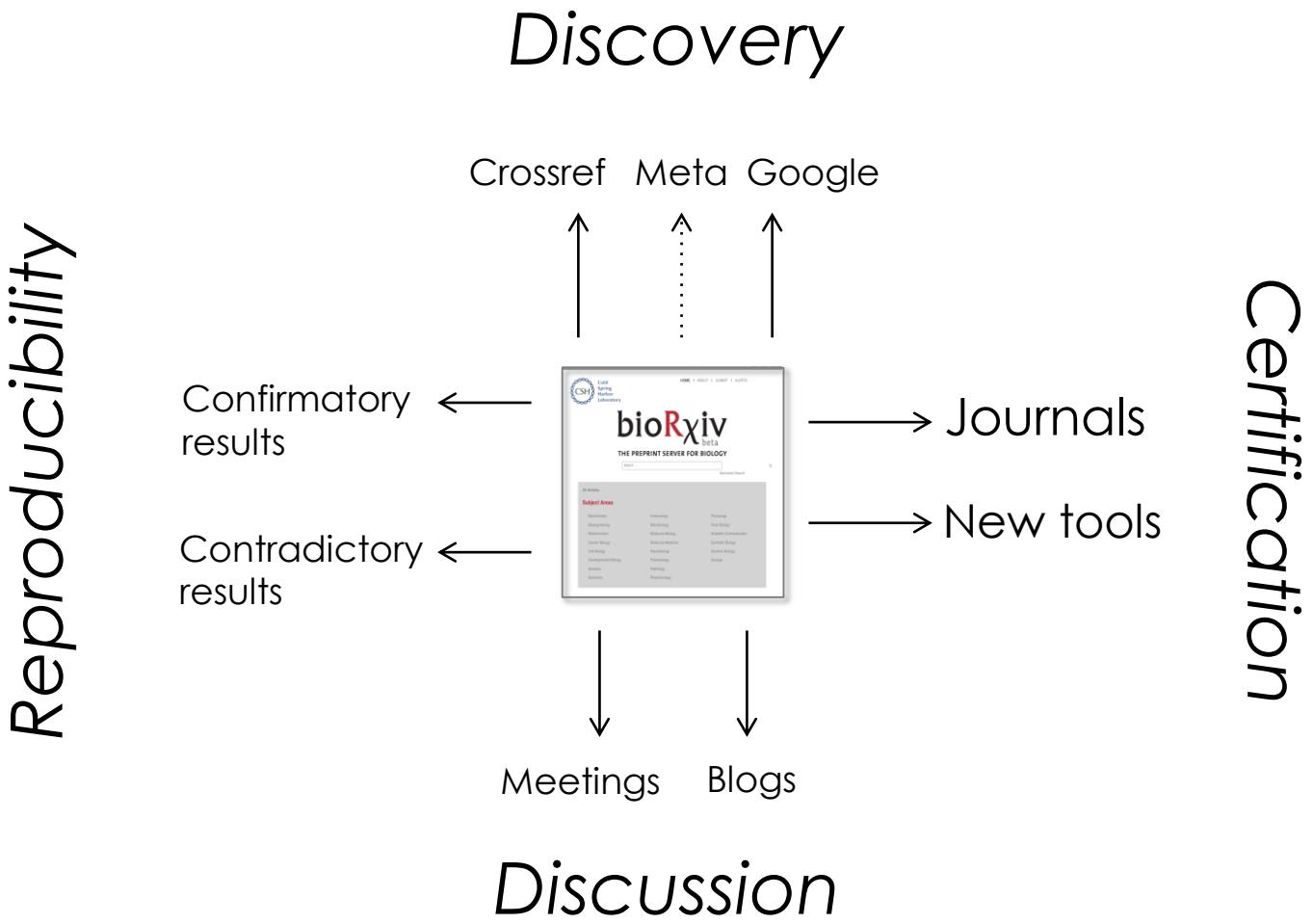
Reluctance to use open licenses:

- No reuse :31% of papers
- CC-BY-ND, NC or both: 51% of papers
- CC-BY: 19% of papers

More education and discussion is needed around the choice of license

Reduction in anxieties about the downside of “giving away content”

bioRxiv - an essential hub



Thank you!

John Inglis

inglis@cshl.edu

@JohnRInglis