

Recommendations to Enhance Rigor and Reproducibility in Biomedical

This manuscript was automatically generated on February 21, 2020.

Authors

- **Jaqueline J. Brito**

 [0000-0002-7158-3253](#) ·  [jaquejbrito](#) ·  [jaquejbrito](#)

Department of Clinical Pharmacy, School of Pharmacy, University of Southern California

- **Jiu Li**

Department of Computational Medicine & Bioinformatics, Medical School, University of Michigan

- **Jason H. Moore**

Department of Biostatistics, Epidemiology, and Informatics, Institute for Biomedical Informatics, University of Pennsylvania

- **Casey S. Greene**

 [0000-0001-8713-9213](#)

Department of Systems Pharmacology and Translational Therapeutics, Perelman School of Medicine, University of Pennsylvania,; Childhood Cancer Data Lab

- **Nicole A. Nogoy**

 [0000-0002-5192-9835](#)

GigaScience

- **Lana X. Garmire**

Department of Computational Medicine & Bioinformatics, Medical School, University of Michigan

- **Serghei Mangul**

 [0000-0003-4770-3443](#) ·  [smangul1](#)

Department of Clinical Pharmacy, School of Pharmacy, University of Southern California

Abstract

Computational methods have reshaped the landscape of modern biology. While the biomedical community is increasingly dependent on computational tools, the mechanisms ensuring open data, open software, and reproducibility are variably enforced by academic institutions, funders and publishers. Publications may describe the software for which source code is unavailable, documentation is incomplete or unmaintained, and analytical source code is missing. Publications that lack this information compromise the role of peer review in evaluating technical strength and scientific contribution. Such flaws also limit any subsequent work that intends to use the described software. We herein provide recommendations to improve reproducibility, transparency, and rigor in computational biology—precisely the values which should be emphasized in foundational life and medical science curricula. Our recommendations for improving software availability, usability, and archival stability aim to foster a sustainable data science ecosystem in biomedicine and life science research.

Keywords: Rigor, reproducible research, installability, archival stability, big data, open science

Introduction

teste

Biomedical informatics has the opportunity to be at the forefront of the biomedical community in practices around open data, open software, and reproducible research. The computational reproduction of previously published results is enabled by publicly releasing all research resources, from raw data to installable packages and source code, in a discoverable and archivally stable manner. However, a lack of strict implementation or enforcement of journal, academic institutions and funder policies for resource sharing has resulted in the loss of crucial research objects for many published studies. Publications lacking data or source code undermine scientific rigor, transparency, and reproducibility [1](#).

An astonishing number of bioinformatics software tools are designed each year to accommodate increasingly bigger, more complex, and more specialized biomedical datasets [2](#). Many of those software tools have limited installability and are hosted on Uniform Resource Locators (URLs) with undetermined archiving practices [3](#). Closed-source software is another issue as researchers may not have access to the source code, which harms the transparency of research by avoiding others to audit methods and results. Here we discuss the pressing need for scientists to improve software availability, usability, and archival stability in computational biology. By following a set of best practices [4](#), scientists can promote rigor and reproducibility, ultimately cultivating a sustainable, thriving research community.

1. Increase computational training opportunities targeted at reproducibility

Biomedical researchers who use computational tools must acquire specific computational skills in order to successfully apply the techniques to a large amount of data. Undergraduate students who lack formal computational training can be taught the skills required to promote reproducibility on specialized courses [5](#). In addition to rigorous class training, advanced undergraduate and graduate students may benefit from short-term intensive workshops aimed at postdoctoral scholars, clinical fellows, and faculty [6](#). Several institutions, including the University of California at Los Angeles, have successfully hosted workshop-based programs for over five years and serve as valuable resources for pedagogy and curricula [7,8](#). Workshops for training researchers to use computational tools usually include hands-on training for implementing analysis tools, such as computational notebook

technologies. Since 1998, Software Carpentry (<https://software-carpentry.org/>) has been holding volunteer-based training courses for researchers with the computational skills required to keep up with the demands of data- and computational-intensive research. In summary, comprehensive computational training programs are the way to prime future computational biologists towards reproducibility work in the data-driven fields of life science and biomedical research (**Figure 1a**).

2. Make all data and metadata open and discoverable

Open source code depends on open and shareable data. Access to the data used to produce important research results is key for auditing the rigor of published studies. Open access to datasets is imperative to build a thriving and sustainable scientific community where all researchers can access and analyze existing data. However in practice, omics data of patients often can not be shared publicly due to patient privacy and/or user agreement standards [9](#), [10](#), [11](#), [12](#), [13](#), [14](#), [15](#), [16](#). While not all data are freely publicly available, many studies provide controlled access, where researchers can access the raw data if they are approved after providing the reason to and signing user agreements. Also, in many cases where raw data are not accessible by the public use, summary data are available. The global data sharing climate in general has shifted towards a positive direction.

In general, the data should also be discoverable via centralized repositories like SRA and GEO and annotated with descriptive metadata to enhance data reuse (**Figure 1b**). When data is shared on centralized repositories in interoperable formats, other researchers can examine and re-analyze the data, challenge existing interpretations, and test new theories. In general, data sharing corresponds to a true spirit of science where each new discovery is built upon previous work - allowing us to “stand on the shoulders of giants”. Discoveries in multiple fields were solely based on shared data (e.g., economics, meteorology, and physics). Sharing data can enable the reproducibility and robustness of science because of the ability to utilize data generated from individual studies to a larger scale. Additionally, secondary analysis is economically sustainable and can be used in countries with limited resources [17](#), [18](#). In general, reusing the data speaks about the quality and importance of generated data and contributed to the impact of the original work.

3. Build and use open-source software

The software provides a foundation for the reproducibility of published biomedical research, defined as the ability to replicate published findings by running the same computational tool on data generated by the study [4](#),[5](#). For this reason, closed-source and proprietary software restricts the reproducibility of biomedical research. First, researchers may not have access to the source code, which limits their ability to audit the results and reviewers are unable to test the reproducibility prior to being published. Second, license restrictions may prohibit the creation of new functionalities that would be released on modified versions of existing tools. For proprietary software, generally, there is also the cost of acquisition and maintenance of licenses, which is not affordable for every laboratory. Reviewers may also not have access to proprietary software, and are thus, unable to fully test reproducibility of results. Therefore, the adoption of standard open source licenses for data and software tools can enhance the rigor and impact of research by allowing others to reproduce published studies.

However, only publicly releasing the code does not guarantee the computational reproducibility of biomedical research. The open code must be well documented with user manuals, and installable in a user-friendly manner. The code used in a published analysis should be hosted on an archivally stable platform such as GitHub (**Figure 1c**). Currently, over a quarter of computational software resources cannot be accessed through the URLs provided in the original publication, suggesting that the repositories are poorly maintained [3](#). Additionally, many bioinformatics tools are too difficult, and some impossible, for a new user to install [3](#). Use of Open Source Initiative license models

(<https://opensource.org/licenses>) allows users to easily use and adapt tools, increasing the sustainability of the biomedical research community. New platforms are also being proposed, such as CODE CHECK (<https://sje30.github.io/codecheck/>), where researchers can, input their code and data, and check that their analysis is reproducible. For verified analyses, CODE CHECK issues a time-stamped “*certificate of reproducible computation*” which can inform the peer reviewer of a paper.

4. Use platforms that enhance reproducibility

In addition to software and datasets, computational biology researchers commonly produce resources such as experiment protocols, workflows, and annotations. Storing and sharing these resources allows other researchers to cite them within a publication, which would increase the reproducibility of a paper and increase the visibility of previously developed methods. The inclusion of citable digital object identifiers (DOIs) also facilitates the discovery of reusable resources as they provide long-term access to published resources. Several innovative platforms designed to promote reproducibility have recently emerged (**Figure 1d**).

5. Make tools and workflows reproducible

To facilitate the reproducibility of tools it is good practice to utilize virtual machines (VMs) and containers. VMs are software pieces that are capable of encapsulating entire operating systems, libraries, codes, and data. For workflows, it is good practice to utilize workflow-specific platforms to enhance reproducibility [19](#), such as Galaxy, Bioconda and workflow standards, such as CWL (Common Workflow Language) (**Table 1**). There are now a myriad of platforms and tools available that make research reproducible - **Table 1** below is a summary of some commonly used. Given the many different tools and platforms available that can promote reproducibility, a research lab should define their own standards on a suite of tools and platforms that support their research practices (**Figure 1e**).

Table 1:

Platform & Type	Use
Reproducible and open methods	Protocols.io An open-source protocol repository, where researchers can manage, share, tweak, optimize and adopt existing methods even after a scientist has left a lab.
RRIDs	Scicrunch.org A platform for curating research resources that enables the user to discover, access, view, and use research objects. Users can register any research object, such as tools, antibodies, animal models etc In turn,. these get issued a RRID (Research Resource ID) which should be cited in the manuscript. The RRID allows other users to easily locate and access the resources.
Annotations	Hypothes.is Is an open-source annotation tool that allows any researcher to annotate any resource on the web, including to create annotations for personal use or as part of conversations with private groups or the general public.
Virtual Machines & Containers	Containers such as Docker (www.docker.com), and Singularity (singularity.lbl.gov) are lightweight solutions compared to VMs as they do not encapsulate the operating system; rather, they rely on the host kernel to run required functions. Both VMs and containers are shared via image files and can be included as supplementary material at certain journals or stored in Zenodo (https://zenodo.org/), Figshare (https://figshare.com/), or other general-purpose archival repositories.

Platform & Type	Use
Reproducible workflows	<p>Galaxy (https://galaxyproject.org/) - a computational platform which allows users to share workflows, histories and wrapped tools in an easy-to-use and open source interface that even people without coding experience can use.,</p> <p>Conda (https://conda.io/) is a powerful open source package and management system that can quickly install, run and update packages and their dependencies.</p> <p>Bioconda (https://bioconda.github.io/) - leveraging Conda, it is a community project and package manager dedicated to the life sciences and biomedical tools.</p> <p>Common workflow language (CWL), (https://www.commonwl.org) - is an open standard used to describe workflows and tools to make them portable and interoperable across different environments e.g. cloud, cluster or high performance computing.</p>
Reproducible documents & Figures	<p>Stencila (https://stenci.la/) is an open source framework for executable documents and living figures (using R scripts). It supports commonly used environments and tools, such as Jupyter notebook, RMarkdown, Python and SQL.</p> <p>Jupyter Notebook (https://jupyter.org/) - allows for the creation of sharing of live code, equations, visualizations and narrative text. The application supports over 40 different programming languages and can be used to leverage big data.</p> <p>MyBinder (https://mybinder.org/) - is an application that collects and 'binds' interactive jupyter notebooks into a Binder repository and can also create a Docker image of that.</p>

6. Make living and reproducible figures and papers

Open data and code is one thing; however over the last 5 years it has been possible to break away from the static presentation of results and produce dynamic/ living figures (**Figure 1f**) - where a reader can alter the parameters and code is actively running [20](#). In principle, the things associated with such figures, include the raw data, and R code to evaluate the data, as well as the functionality of the paper to call the data using software to compute an evaluation. One such example is Stenci.la - a platform for executable documents and living figures and Jupyter Notebook (**Table 1**).

7. Publish with journals that promote reproducibility

Journals have various publishing standards. Researchers may elect to publish in journals that encourage best practices (e.g., adopting the FAIR principles [21](#)) that aim to increase the impact of their work (**Figure 1g**). A group of stakeholders from academia and industry defined a set of principles stating that research data should be Findable, Accessible, Interoperable, and Reusable (FAIR) [21](#). To ensure reproducibility, many journals now require that biomedical data generated by a published study be shared when the paper is released. For instance, *GigaScience* (gigasciencejournal.com) has been promoting reproducibility of analyses since 2012 by mandating open data that follow the FAIR principles and mandates source code with an OSI approved license. During peer review, *GigaScience* also makes all supporting data and code available for Reviewers - asking them to test for reproducibility; and papers presenting VMs, containers, Jupyter notebooks or packaged workflows makes this task easier. Other journals, such as *Biostatistics*, have implemented badges for articles

with data and code sharing [22](#)). *eLife* published their first example of a dynamic and computationally reproducible article - which enables data and analysis to be fully reproducible by the reader [23](#). It is a code-based publication (using the Stencila platform, **Table 1**), challenging the traditional static representation of results using PDF or HTML formats.

8. Support reusable resources

Successfully implementing and widely distributing software tools developed in academia involves unique challenges when compared to the industry. In academia, software tools are developed by small groups comprised of graduate or postdoctoral scholars, who have fairly fast turn-over rates of 2-5 years [24](#). These groups are less professionally trained when compared to software development groups in the industry, where holistic teams of specialists support the long-term maintenance of projects. In order to enhance the quality and reuse of open software, professionally trained software engineers should be hired to partner with the students and postdocs. To make this happen, funding agencies need better mechanisms of acknowledging and incentivizing funding earmarked for critical bioinformatics infrastructure (**Figure 1h**). In addition, funders should recognize the rigor of software development, rather than just considering 'novelty'-based conventional criteria of research. The availability of well-resourced grant mechanisms to convert minimum viable products produced by trainees into reliable software could enhance the impact of research-grade software on the community [10](#). With the growing number of biomedical datasets open for reuse in the public domain, It is inspiring to see the encouragement and acknowledgement of data reuse and secondary analysis with the research Parasite Awards [25](#). The annual Parasite Awards recognizes the exceptional contributions for rigorous secondary analysis of data with two award tracks - Junior parasite and continued research parasitism. More incentives, such as this, are required for software reuse.

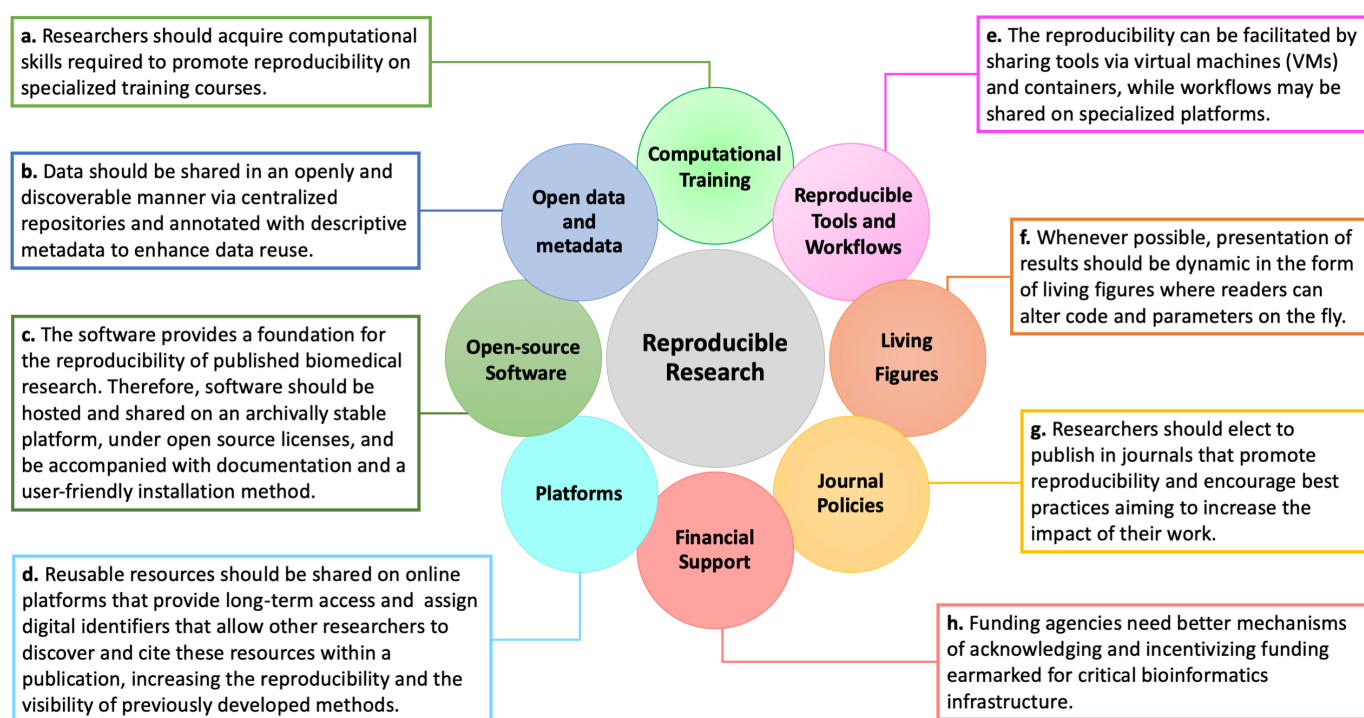


Figure 1. Recommendations to improve reproducibility and rigor of research.

Conclusions

We outlined key recommendations to improve the rigor of biomedical studies and foster reproducibility in computational biology. The infrastructure required to systematically adopt best

practices for reproducibility of biomedical research is largely in place. The remaining challenge to the systemic promotion of reproducibility is that incentives are not currently aligned to support good practices. Instead, current efforts rely on individual researchers electing to follow best practices, often at their own time and expense. We believe it is time for a fundamental cultural shift in the scientific community: rigor and reproducibility should become primary concerns in the criteria and decision-making process of designing studies, funding research, and writing and publishing results. Successful systematic adoption of best practices will require the buy-in of multiple stakeholders in the scientific communities, from publishers, academic institutions, funding agencies, and stakeholders. This increases the lifetime and value of published research as resources naturally become reusable, testable, and discoverable. Community-wide adoption of best practices for reproducibility is critical to realizing the full potential of fast-paced, collaborative analyses of large datasets in the biomedical and life sciences.

Abbreviations

CWL, Common Workflow Language; DOI, Digital Object identifier; FAIR, Findable, Accessible, Interoperable, and Reusable; OSI, Open Source Initiative; VM, Virtual Machine.

Competing interests

The authors declare that they have no competing interests.

Funding

CSG was supported by grants from the NIH (R01HG010067 and R01CA237170), the Gordon and Betty Moore Foundation (GBMF4552), the Chan Zuckerberg Initiative Donor Advised Fund of the Silicon Valley Community Foundation (2018-182718), and Alex's Lemonade Stand Foundation (CCDL). LXG is supported by grants K01ES025434 awarded by NIEHS through funds provided by the trans-NIH Big Data to Knowledge (BD2K) initiative (<http://datascience.nih.gov/bd2k>), R01 LM012373 and R01 LM012907 awarded by NLM, and R01 HD084633 awarded by NICHD. The funding bodies had no role in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript.

Authors' contributions

Author	Contribution
JJB	Writing - Original Draft Preparation, review and editing.
JL	Writing - review and editing.
CSG	Writing - review and editing.
JM	Writing - review and editing.
NN	Conceptualization and structure of the manuscript; Writing - review and editing; Visualization - creation of Figures and Table.
LG	Conceptualizing the project; Writing - review and editing.
SM	Conceptualization and structure of the manuscript; Writing - review and editing.

Acknowledgments

We thank Dr. Lana Martin for discussions and helpful comments on the manuscript.

Authors' Information

NAN is an Editor at *GigaScience* and is an open science advocate with over 7 years experience in publishing reproducible research.

References

- 1. An empirical analysis of journal policy effectiveness for computational reproducibility**
Victoria Stodden, Jennifer Seiler, Zhaokun Ma
Proceedings of the National Academy of Sciences (2018-03-12) <https://doi.org/gc8gkw>
DOI: [10.1073/pnas.1708290115](https://doi.org/10.1073/pnas.1708290115) · PMID: [29531050](https://pubmed.ncbi.nlm.nih.gov/29531050/) · PMCID: [PMC5856507](https://pubmed.ncbi.nlm.nih.gov/PMC5856507/)
- 2. Bioinformatics programs are 31-fold over-represented among the highest impact scientific papers of the past two decades.**
Jonathan D Wren
Bioinformatics (Oxford, England) (2016-05-05) <https://www.ncbi.nlm.nih.gov/pubmed/27153671>
DOI: [10.1093/bioinformatics/btw284](https://doi.org/10.1093/bioinformatics/btw284) · PMID: [27153671](https://pubmed.ncbi.nlm.nih.gov/27153671/)
- 3. Improving the usability and archival stability of bioinformatics software**
Serghei Mangul, Lana S. Martin, Eleazar Eskin, Ran Blekhan
Genome Biology (2019-02-27) <https://doi.org/ggc6j6>
DOI: [10.1186/s13059-019-1649-8](https://doi.org/10.1186/s13059-019-1649-8) · PMID: [30813962](https://pubmed.ncbi.nlm.nih.gov/30813962/) · PMCID: [PMC6391762](https://pubmed.ncbi.nlm.nih.gov/PMC6391762/)
- 4. Tools and techniques for computational reproducibility**
Stephen R. Piccolo, Michael B. Frampton
GigaScience (2016-07-11) <https://doi.org/gfs3cq>
DOI: [10.1186/s13742-016-0135-4](https://doi.org/10.1186/s13742-016-0135-4) · PMID: [27401684](https://pubmed.ncbi.nlm.nih.gov/27401684/) · PMCID: [PMC4940747](https://pubmed.ncbi.nlm.nih.gov/PMC4940747/)
- 5. Strategies for Building Computing Skills To Support Microbiome Analysis: a Five-Year Perspective from the EDAMAME Workshop**
Ashley Shade, Taylor K. Dunivin, Jinlyung Choi, Tracy K. Teal, Adina C. Howe
mSystems (2019-08-20) <https://doi.org/ggkt8k>
DOI: [10.1128/msystems.00297-19](https://doi.org/10.1128/msystems.00297-19) · PMID: [31431509](https://pubmed.ncbi.nlm.nih.gov/31431509/) · PMCID: [PMC6702294](https://pubmed.ncbi.nlm.nih.gov/PMC6702294/)
- 6. Making Workshops Work: Insights from EDAMAME**
Jaclyn N. Taroni
mSystems (2019-08-20) <https://doi.org/ggkt8m>
DOI: [10.1128/msystems.00467-19](https://doi.org/10.1128/msystems.00467-19) · PMID: [31431510](https://pubmed.ncbi.nlm.nih.gov/31431510/) · PMCID: [PMC6702295](https://pubmed.ncbi.nlm.nih.gov/PMC6702295/)
- 7. Addressing the Digital Divide in Contemporary Biology: Lessons from Teaching UNIX.**
Serghei Mangul, Lana S Martin, Alexander Hoffmann, Matteo Pellegrini, Eleazar Eskin
Trends in biotechnology (2017-07-15) <https://www.ncbi.nlm.nih.gov/pubmed/28720283>
DOI: [10.1016/j.tibtech.2017.06.007](https://doi.org/10.1016/j.tibtech.2017.06.007) · PMID: [28720283](https://pubmed.ncbi.nlm.nih.gov/28720283/)
- 8. greenelab/GCB535**
Greene Laboratory
(2019-04-19) <https://github.com/greenelab/GCB535>
- 9. Data Access in Genomics**<https://www.nature.com/collections/diadjciaj>

10. Data sharing and the future of science

Nature Communications

(2018-07-19) <https://doi.org/cr99>

DOI: [10.1038/s41467-018-05227-z](https://doi.org/10.1038/s41467-018-05227-z) · PMID: [30026584](https://pubmed.ncbi.nlm.nih.gov/30026584/) · PMCID: [PMC6053389](https://pubmed.ncbi.nlm.nih.gov/PMC6053389/)

11. No impact without data access

Nature Genetics

(2015-06-26) <https://doi.org/ggkt8q>

DOI: [10.1038/ng.3351](https://doi.org/10.1038/ng.3351) · PMID: [26111506](https://pubmed.ncbi.nlm.nih.gov/26111506/)

12. Analysis commons, a team approach to discovery in a big-data environment for genetic epidemiology

Jennifer A BrodyAlanna C Morrison, Joshua C Bis, Jeffrey R O'Connell, Michael R Brown, Jennifer E Huffman, Darren C Ames, Andrew Carroll, Matthew P Conomos, ...

Nature Genetics (2017-10-27) <https://doi.org/gchz7j>

DOI: [10.1038/ng.3968](https://doi.org/10.1038/ng.3968) · PMID: [29074945](https://pubmed.ncbi.nlm.nih.gov/29074945/) · PMCID: [PMC5720686](https://pubmed.ncbi.nlm.nih.gov/PMC5720686/)

13. Analysis of five years of controlled access and data sharing compliance at the International Cancer Genome Consortium

Nature Genetics (2016-02-24) <https://doi.org/ggkt8r>

DOI: [10.1038/ng.3499](https://doi.org/10.1038/ng.3499) · PMID: [26906679](https://pubmed.ncbi.nlm.nih.gov/26906679/)

14. DNA.Land is a framework to collect genomes and phenomes in the era of abundant genetic information

Jie Yuan, Assaf Gordon, Daniel Speyer, Richard Aufrichtig, Dina Zielinski, Joseph Pickrell, Yaniv Erlich

Nature Genetics (2018-01-26) <https://doi.org/gdcc77>

DOI: [10.1038/s41588-017-0021-8](https://doi.org/10.1038/s41588-017-0021-8) · PMID: [29374253](https://pubmed.ncbi.nlm.nih.gov/29374253/)

15. UK Biobank debuts as a powerful resource for genomic research.

Teri A Manolio

Nature medicine (2018-12) <https://www.ncbi.nlm.nih.gov/pubmed/30510254>

DOI: [10.1038/s41591-018-0276-3](https://doi.org/10.1038/s41591-018-0276-3) · PMID: [30510254](https://pubmed.ncbi.nlm.nih.gov/30510254/)

16. Using and producing publicly available genomic data to accelerate discovery in nephrology

Matthew G. Sampson, Hyun Min Kang

Nature Reviews Nephrology (2019-06-10) <https://doi.org/ggkt8s>

DOI: [10.1038/s41581-019-0166-z](https://doi.org/10.1038/s41581-019-0166-z) · PMID: [31182850](https://pubmed.ncbi.nlm.nih.gov/31182850/)

17. Using bioinformatics training to boost research capacities in resource-limited regions

Serghei Mangul, Lana Martin, Ben Langmead, Javier Sanchez Galan, Ian Toma, Pavel Pevzner, Eleazar Eskin

PeerJ (2018-12-06) <https://doi.org/ggkt8t>

DOI: [10.7287/peerj.preprints.27415v1](https://doi.org/10.7287/peerj.preprints.27415v1)

18. How bioinformatics and open data can boost basic science in countries and universities with limited resources

Serghei Mangul, Lana S. Martin, Ben Langmead, Javier E. Sanchez-Galan, Ian Toma, Fereydoun Hormozdiari, Pavel Pevzner, Eleazar Eskin

Nature Biotechnology (2019-03) <https://doi.org/gfwnsm>

DOI: [10.1038/s41587-019-0053-y](https://doi.org/10.1038/s41587-019-0053-y) · PMID: [30833765](https://pubmed.ncbi.nlm.nih.gov/30833765/)

19. Practical Computational Reproducibility in the Life Sciences

Björn Grüning, John Chilton, Johannes Köster, Ryan Dale, Nicola Soranzo, Marius van den Beek, Jeremy Goecks, Rolf Backofen, Anton Nekrutenko, James Taylor

Cell Systems (2018-06) <https://doi.org/ggdv3z>
DOI: [10.1016/j.cels.2018.03.014](https://doi.org/10.1016/j.cels.2018.03.014) · PMID: [29953862](https://pubmed.ncbi.nlm.nih.gov/29953862/) · PMCID: [PMC6263957](https://pubmed.ncbi.nlm.nih.gov/PMC6263957/)

20. Referee report. For: Sub-strains of *Drosophila* Canton-S differ markedly in their locomotor behavior [v1; approved 1, <http://f1000r.es/3is>]

Gregg Roman

F1000Research (2014) <https://doi.org/ggkt8w>

DOI: [10.5256/f1000research.4564.r5630](https://doi.org/10.5256/f1000research.4564.r5630)

21. The FAIR Guiding Principles for scientific data management and stewardship

Mark D. Wilkinson, Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, Jan-Willem Boiten, Luiz Bonino da Silva Santos, Philip E. Bourne, ... Barend Mons

Scientific Data (2016-03-15) <https://doi.org/bdd4>

DOI: [10.1038/sdata.2016.18](https://doi.org/10.1038/sdata.2016.18) · PMID: [26978244](https://pubmed.ncbi.nlm.nih.gov/26978244/) · PMCID: [PMC4792175](https://pubmed.ncbi.nlm.nih.gov/PMC4792175/)

22. Badges for sharing data and code at Biostatistics: an observational study

Anisa Rowhani-Farid, Adrian G. Barnett

F1000Research (2018-03-07) <https://doi.org/gdsnhw>

DOI: [10.12688/f1000research.13477.2](https://doi.org/10.12688/f1000research.13477.2) · PMID: [29862016](https://pubmed.ncbi.nlm.nih.gov/29862016/) · PMCID: [PMC5843843](https://pubmed.ncbi.nlm.nih.gov/PMC5843843/)

23. Replication Study: Transcriptional amplification in tumor cells with elevated c-Myc

L Michelle Lewis, Meredith C Edwards, Zachary R Meyers, C Conover Talbot Jr, Haiping Hao, David Blum, Elizabeth Iorns, Rachel Tsui, Alexandria Denis, Nicole Perfito, ...

eLife (2018-01-09) <https://doi.org/gf9h4w>

DOI: [10.7554/elife.30274](https://doi.org/10.7554/elife.30274) · PMID: [29313490](https://pubmed.ncbi.nlm.nih.gov/29313490/) · PMCID: [PMC5760205](https://pubmed.ncbi.nlm.nih.gov/PMC5760205/)

24. The anatomy of successful computational biology software

Stephen Altschul, Barry Demchak, Richard Durbin, Robert Gentleman, Martin Krzywinski, Heng Li, Anton Nekrutenko, James Robinson, Wayne Rasband, James Taylor, Cole Trapnell

Nature Biotechnology (2013-10) <https://doi.org/gf8nr4>

DOI: [10.1038/nbt.2721](https://doi.org/10.1038/nbt.2721) · PMID: [24104757](https://pubmed.ncbi.nlm.nih.gov/24104757/) · PMCID: [PMC4166496](https://pubmed.ncbi.nlm.nih.gov/PMC4166496/)

25. Celebrating parasites.

Casey S Greene, Lana X Garmire, Jack A Gilbert, Marylyn D Ritchie, Lawrence E Hunter

Nature genetics (2017-03-30) <https://www.ncbi.nlm.nih.gov/pubmed/28358134>

DOI: [10.1038/ng.3830](https://doi.org/10.1038/ng.3830) · PMID: [28358134](https://pubmed.ncbi.nlm.nih.gov/28358134/) · PMCID: [PMC5710834](https://pubmed.ncbi.nlm.nih.gov/PMC5710834/)