

**National Human Genome Research Institute**  
**Virtual Meeting to Discuss Genomic Data Sharing with Journal Editors**

November 30, 2020

Meeting Summary

## Background

On November 30<sup>th</sup>, 2020, The National Human Genome Research Institute (NHGRI) convened a virtual meeting for journal editors and NHGRI/NIH staff to discuss genomic data sharing. The goal of this meeting was to learn more about the issues faced by scientific publishers that publish genomics papers (“Journals”); and to develop ideas to improve data sharing policy, its implementation, and the practice of leveraging data standards to optimize the utility of genomic data shared by researchers for secondary research purposes.

The [NHGRI Strategic Vision](#)<sup>1</sup>, published in October 2020, articulates data sharing as a guiding principle and value of NHGRI and the broader genomics research community. Thus, as the Institute enters its fourth decade of supporting research at the Forefront of Genomics, the meeting was organized to gather information on how to improve genomic data sharing for the various stakeholders involved in the endeavor.

NHGRI asked meeting participants to respond to several questions before the meeting to help inform priority areas for the discussion:

1. What are challenges in the genomic data sharing arena that you are currently facing? What do you worry about or expect might come up related to genomic data sharing?
2. Are there any examples of successes? What actions led to these successes?
3. Could you share with us an example of when things did not go smoothly?
4. Are there ways that funders and journals can work together to improve the “FAIRness” of genomic data sharing?
5. Please share a link to your current genomic data sharing policies, if possible.

*See **Appendix 1** for a listing of Journal Data Policies.*

## Opening Remarks

*Eric Green, Director, NHGRI*

Dr. Green welcomed the group and began by summarizing the four major sections of the recently published NHGRI 2020 Strategic Vision, which is the product of over two years of planning and was informed by extensive outreach by the Institute. The 2020 Vision is organized into four distinct areas: 1) Guiding principles and values for human genomics; 2) Sustaining and improving a robust foundation for genomics; 3) Breaking down barriers that impede progress in genomics; and, 4) Compelling genomics research projects in biomedicine. The spirit of this meeting is aligned with the Vision as it aims to explore important issues at the Forefront of Genomics.

Genomic data sharing is in fact one of the *guiding principles and values* of the Vision, stated as, “**Adhere to the highest expectations and requirements related to open science, responsible data sharing, and**

---

<sup>1</sup> Green, E. D., Gunter, C., Biesecker, L. G., Di Francesco, V., Easter, C. L., Feingold, E. A., ... & Manolio, T. A. (2020). Strategic vision for improving human health at The Forefront of Genomics. *Nature*, 586(7831), 683-692.

**rigor and reproducibility in genomics research** — the genomics enterprise has a well-respected history of leading in these areas, and that commitment must be built upon and continually reaffirmed.”

NHGRI, serving in a leadership role, acts as a responsible steward for this and the other principles and values; this role includes the duty to constantly assess genomic data sharing policies and implementation to ensure the pursuit is optimally tuned to ensure open science and reduce data-access burdens to advance research, including the use of optimally balanced and ethically sound approaches for respecting participant preferences and consent as well as engaging communities.. This meeting was convened in recognition that current genomic data sharing is very good, but improvements can be made particularly in the rapidly changing environment of scientific publishing. NHGRI invited editors primarily from genetics and genomics publications, as such Journals play a critical role in ensuring and evaluating data sharing, to seek their help with identifying critical issues and barriers and brainstorming possible solutions.

Dr. Green thanked everyone for attending, and particularly those who took time to answer the pre-meeting questions to help inform this meeting. He also thanked the meeting organizers and those helping behind the scenes with the logistics for the meeting.

*Carolyn Hutter, Division Director, Division of Genome Sciences, NHGRI*

Dr. Hutter set the stage for the meeting with an overview of NIH and NHGRI genomic (and non-genomic) data sharing policies and practices. She also emphasized that this is not a one-time discussion; NHGRI sees this work as an ongoing partnership and this meeting as a springboard to focus on specific areas in the future.

The timeliness and importance of this topic is reflected by the emphasis on data sharing in the Strategic Vision. NHGRI is interested in evolving and building upon previous activities and successes.

This discussion is meant to focus on what can be done by Funders and Journals, however, anything these two groups might pursue is done in the context of other players (e.g., standards-generating bodies, such as GA4GH; participants in genomics studies; research institutions; IRBs; data generators; end-point data users; and repositories). Some things can be driven by Funders and Journals, whereas others may require expanding partnerships with other stakeholders to make additional improvements.

NIH data sharing policies have evolved over time. Dr. Hutter highlighted a few, including the 2008 [NIH Genome-wide Association Studies \(GWAS\) Policy](#), which later evolved into 2014 [NIH Genomic Data Sharing \(GDS\) Policy](#). She also highlighted the newly announced 2020 [NIH Data Management and Sharing Policy](#). These policies were also influenced somewhat by a number of other activities coming out of the White House and HHS. Importantly, each policy also has aspects that are open areas for interpretation and implementation.

In the U.S. the NIH GDS Policy sets forth an expectation of broad and responsible sharing of research genomic data by outlining requirements for *submitters* of data (generating a genomic data sharing plan, with specific expectations for sharing human and non-human data on certain timelines). It also provides a framework for data *users* access to genomic data, with additional expectations for those using large-scale human genomic data. Finally, it describes Intellectual Property considerations for genomic data.

NHGRI, as a leader in genomic data sharing, [builds upon the foundation laid by the NIH GDS Policy](#) in its implementation and interpretation of the policy in a number of ways. NHGRI encourages sharing of all genomic data and data types. NHGRI expects all data to be shared on similar timelines, regardless of whether data are derived from humans or other organisms. The Institute also encourages human studies to use sources with consent for General Research Use through controlled access or sources with consent for unrestricted access. Finally, NHGRI is implementing a new expectation (in January 2021) to move toward consistency in the expectation for *explicit consent* for broad data sharing and future research use across all NHGRI-funded studies.

The new NIH Data Management and Sharing (DMS) Policy applies to all research funded or conducted by NIH that results in the generation of scientific data. The effective date of the new Policy is approximately two years after its publication—January 25, 2023. The Policy requires that all researchers submit a DMS Plan *when applying for funding*. As outlined in supplemental guidance, [Plans should describe](#) how the data will be generated, how it will be managed, and which of the data and accompanying metadata will be shared. Compliance with the Plan is the second requirement of this Policy. Because data management and sharing will require additional funds, NIH also provided [guidance on allowable costs](#). A third supplemental guidance document also provided [information about which repositories should be used](#), and what criteria should be evaluated if an NIH-established resource is not readily available.

The DMS Policy recognizes metadata sharing, as well. The increased focus of the sharing of metadata aligns with a growing NIH and NHGRI focus on the [FAIR Principles](#) (FAIR = Findable, Accessible, Interoperable, and Reusable). Metadata are key to achieving the FAIR Principles, and thus are key to new NIH DMS Policy.

In parallel to the NIH's DMS Policy expectation of metadata sharing, NHGRI also plans to increase the emphasis on the sharing and availability of metadata and phenotypic data through a Guide Notice. Though the emphasis will not be limited to sharing under the NIH GDS Policy, the movement at NHGRI aims to better state NHGRI's interpretation of the metadata and phenotypic data that should be shared to meet the NIH GDS Policy's expectation for the sharing of 'relevant associated data.' A recommendation from Mark Johnston (*Genetics*) on this Guide Notice (under the auspices of the NHGRI Genomic Data Science Working Group) led to the idea to discuss this topic with Journal Editors.

Dr. Hutter then shifted to provide several examples where NHGRI feels there has been a lot of success to date to help illustrate the goals and benefits of broad genomic data sharing. First, many large genomics resource building projects (e.g., those that were designed with the mission of broad data sharing such as ENCODE, HubMap, Human Cell Atlas, have both [utilized and informed data standards](#) for genomic data and metadata. ENCODE, for example, uses a highly standardized and relational metadata model. This allows for citation of specific accessions, improves searches across the ENCODE portal, and facilitates interoperability with other datasets. ENCODE recently published information about the strong uptake of data reuse (community publications are outpacing the number of publications from the ENCODE consortium investigators themselves). This is likely influenced by the commitment to data sharing and the organization of the data through ontologies. Dr. Hutter pointed out that ENCODE is also very good at engaging with other consortia when developing standards (working with IHEC and other groups through GA4GH to work on standards for sharing epigenomic data), and their clear policies for sharing software and analysis methods, which allows for integration with other projects that can use those methods and tools.

Another example of successful data sharing is the GWAS catalog—an NHGRI-EBI collaboration. The GWAS catalog captures GWAS results and allows for curation of GWAS publications. The project has expanded beyond gathering information from publications to capture pre-publication data. In addition, they now also provide summary statistics across data, rather than just top hits, which enables metaanalysis and other follow-up studies. Excitingly, the project also plans to expand to new data types and formats for sequencing studies, to create a PRS catalog, and more. In June 2020, the program held a workshop to discuss needs for moving forward and how to ensure the catalog is maximally useful for all stakeholders. The group defined the importance of recognizing the GWAS catalog as a central resource for GWAS data sharing, and proposed improvements to the methods for submitting summary statistics and its associated metadata. In addition, the stakeholders suggested additional guidance for mitigating potential risks and what is appropriate to share. In order for GWAS data sharing to be maximally useful, it is important to ensure versioning of data to enable people to identify the most recent data, linkage to other relevant resources (e.g., dbGaP), and summary statistics shared in a FAIR manner.

Dr. Hutter’s last example was the most clinically oriented effort. The ClinVar/ClinGen partnership has enabled the sharing of over 1.35 million sequence variant interpretations. This data comes from clinical laboratory testing and interpretation of the variants. Making these interpretations openly available allows comparison, resolution of discrepancies, and helps to inform better care for patients. One of the lessons that emerged was that recognizing data submitters is great way to motivate continued and increased data sharing (ClinGen has a [“recognized lab list”](#) that requires certain criteria be met for inclusion).

Though there have been successes, this meeting was pulled together to discuss areas that require improvement. From an NHGRI perspective, some initial questions to consider included:

- How do we improve the identification of appropriate repositories and ensure that data are deposited correctly (i.e., QC and adequate metadata)?
- How can we come to a shared community understanding of consistent metadata and metadata standards?
- How do we expand our thoughts on data sharing to include standardization and dissemination of code pipelines and analytical tools?
- How do we consider the changing landscape—cloud computing, emerging technologies—of genomic sharing (in the global context, as well)?
- How do we also think about modernizing the approaches taken for data management and sharing to facilitate efficiency of sharing for submitters and accessors?

*Chris Gunter, Senior Advisory to the NHGRI Director on Genomics Engagement, NHGRI*

As a guide for the open discussion portion of the meeting, Dr. Gunter provided a summary of what NHGRI learned from a survey of attending journal editors, received in advance of the meeting, and then led a discussion along with Dr. Veronique Kiermer, Chief Scientific Officer for PLOS.

First, challenges and worries were organized into several themes:

# 1. Challenges and Worries

A C G  
C G T  
A C G

## Consistency:

- Country-specific barriers
- Consistent review across editors
- Private sector wishes to determine access
- "Rich get richer," where only privileged can access data

## Delays / Tracking Data Sharing:

- Submission challenges can delay establishment of an accession number
- Controlled-access datasets can not be evaluated in-depth by reviewers
- Authors may have to go back and redo analyses to remove data that don't meet journal/repository standards

## Differing or Missing Standards:

- Differing standards for human datasets
- Lack of standards for sharing code and/or reproducible pipelines for data analyses
- "Asks" of reviewers are too big, which can allow manipulation or fraud to go undetected
- Linking genotype and clinical phenotype data increases risk of patient identifiability

## Enforcement:

- Editorial staff may have to go multiple rounds with authors to get them to make data available
- Changes in repositories over time: non-human genomic data previously deposited/publicly available in dbVar now harder to find/obtain.

## Lack of Information About Permission to Share:

- Particularly in the clinical space, authors are sometimes unsure about whether samples were consented for broad data sharing
- Unclear whether sharing summary statistics of an unpublished dataset is sufficient if the author is not the primary steward of the dataset



3

Two journal representatives provided examples of successes and best practices. First, Dr. Jan Higgins of *Genetics in Medicine*, provided the example of a pilot project pursued by two journals that determined requiring authors to verify that variants comply with the Human Genome Variation Society standards [was a reasonable first step](#) towards standardizing the worldwide inventory of human variation.<sup>2</sup> Next, Dr. Rabia Begum of *Genome Medicine* offered that reluctance from authors to sharing data can be overcome by emphasizing reasons to share data (e.g., re-iterate the importance of reproducibility of the conclusions drawn and for their work to be of interest; emphasize that the data can be de-identified, and explain how they may do so; where appropriate, suggest controlled-access as a possibility for data sharing, if consented appropriately).

Next, editors reported a few case studies of difficult situations they have had to navigate. First, some reported challenges with deciding whether to publish papers based on "proprietary" datasets. Sometimes, Journals take the stance that there *are no exceptions*; others weigh the importance of the findings against the drawback that the primary data cannot be shared. Almost all respondents brought up case studies of challenges they have faced with authors based in other countries who cited country-specific reasons for not submitting their data for broader sharing. The last use case that was a problem of awareness and education about data sharing: authors often do not realize they need to consider whether the data were consented for data sharing, and that sharing of data is necessary for compliance with a journal's data policy.

Lastly, Dr. Gunter summarized the ideas that were put forth in the responses for areas where Funders, such as NHGRI, and all the Journals could collaborate on improvements/solutions:

<sup>2</sup> Higgins, J., Dagleish, R., den Dunnen, J. T., Barsh, G., Freeman, P. J., Cooper, D. N., ... & Cutting, G. R. (2021). Verifying nomenclature of DNA variants in submitted manuscripts: Guidance for journals. *Human mutation*, 42(1), 3-7.

Funders	Journals	Both
<ul style="list-style-type: none"> <li>• Data sharing and management plans should be required at the outset of a study, to save time at the publication stage</li> <li>• Support for upload, download, and analysis at reasonable cost for less privileged scientists (while preserving privacy and security)</li> <li>• Input on appropriate balance of sharing “raw” and “processed” data</li> <li>• Infrastructure and incentives for peer review of deposited datasets</li> </ul>	<ul style="list-style-type: none"> <li>• Have a robust, and enforced, data sharing policy</li> <li>• Encourage data citations to support findability of the data</li> <li>• Continuously improve submission systems to make data and paper co-submission easier for authors, and checking easier for editors &amp; staff</li> <li>• Update guides to authors as new policies come out</li> <li>• Ongoing discussion with community on reasonable analysis/reanalysis expectations for peer reviewers</li> </ul>	<ul style="list-style-type: none"> <li>• Endorse genomics standards, recommended file types, best practices for reporting computational pipelines, and nomenclature (and support for evolution over time)</li> <li>• Support an international code of conduct for genomic data sharing (<i>Nature</i> paper – <a href="https://doi.org/10.1038/d41586-020-00082-9">https://doi.org/10.1038/d41586-020-00082-9</a>)</li> <li>• Outline a set of standard, FAIR, supported repositories dependent on the data type and restrictions</li> <li>• Researcher education on FAIR data sharing, including other fields</li> </ul>

**Meeting Findings and Proposed Solutions**

**Finding 1:** Sharing of clinically sequenced genomic data (i.e., data that were generated by a testing laboratory or by a healthcare facility for clinical reasons) can be particularly challenging for researchers and Journals. Issues of data *ownership* and *consent* can impede the genomic data sharing process. For small studies and/or rare disease research, there are also patient confidentiality concerns. Additionally, it is challenging to identify/choose consistent data and metadata standards for clinically derived data. Finally, legacy samples often lack the specific consent that is required today for genomic data sharing today.

**Potential Solutions:**

- There are categories of data that can usually always be shared; these should be shared by all authors. For example, summary-level data or data that were already shared in the text of the publication should be deposited in a repository for improved “findability.”
- Both Funders and Journals should identify, and describe up-front, which data standards clinical researchers should use; and engage with the clinical community to fill in the gaps where standards are missing.
- Funders should consider supporting projects to develop tools for researchers to generate/transform metadata into standardized formats.

**Finding 2:** Journals engage in case-by-case evaluations of data sharing that can be time-consuming and inconsistent.

**Potential Solutions:**

- An evaluation of plans for data management and sharing by NIH at the grant application stage (upon implementation of the new NIH Data Management and Sharing Policy in 2023) will be helpful to Journals as it will alleviate some of the burden currently felt by Journals to educate researchers on various aspects of genomic data sharing.
  - o Note: this does not cover all authors, however, especially those from other countries.

**Finding 3:** There is no clear baseline requirement for genomic data sharing that applies across all jurisdictions, funders, and Journals.

**Potential Solutions:**

- Funders and Journals should work with other stakeholders to define a minimum set of expectations (i.e., “[COPE \(Committee on Publication Ethics\) Guidelines](#)” for genomic data sharing) that most/all authors can abide by to increase consistency of expectations.
  - o Include other funders in this discussion.
  - o Define minimal metadata sharing expectations as well.
  - o Engage with the UK’s Research Council/Wellcome Trust/COPE.
  - o Articulate the minimum viable product as well as ‘extras’ that would be nice to have.
  - o Example: *HGG Advances* is piloting a mechanism to eventually require the submission of variant data accompanied by metadata and phenotypic data to open access databases. This information is already in the publication; therefore, consent for sharing is not an issue.

**Finding 4:** Authors need education about and resources for genomic data sharing:

- Guidance on how to share clinically derived data (see issues raised in Finding 1).
- Guidance on which data are sensitive and which data types can be broadly shared.
- Education on the importance of using data and metadata standards, and guidance on which standards to use.

**Potential Solutions:**

- Funders and Journals should provide researchers with education about which data they can share broadly and which data types require more protection.
- Funders should develop, potentially in partnership with Journals, clear educational materials/sessions for researchers on genomic data sharing, particularly about the importance of using data and metadata standards.
- Standards take a long time to develop and a lot of discussion before they are finalized. In addition, it takes time for the word to be disseminated that a standard should be adopted and used. Once a standard has been developed and is widely utilized, it would be helpful for Funders to support the development of tools to help researchers to easily generate standardized data, metadata, and phenotypic data.

**Finding 5:** Journals are currently having to track and consider international differences in researchers’ capacity to share data due to national or regional regulations. Some countries take a proprietary view to the health data from their citizens and thus restrict sharing outside of the country. The General Data



Protection Regulation (GDPR) is causing hesitancy/uncertainty about the sharing of biomedical research data outside of certain jurisdictional borders.

Other data sharing limitations stem from regions/nations where there are concerns about abuse (e.g., tribal nations). Editors also raised the issue of having to look out for cases of data colonialism/appropriation. Differences in culture around informed consent also can create consent issues from certain regions.

**Potential Solutions:**

- Cloud-based sharing may help with some jurisdictional issues.
  - o Note: Certain countries may require that cloud platforms must be located physically within the country's borders to meet the jurisdictional requirements.
- As with the solution under "Finding 3" above, some standard, minimal expectations that can work across most countries would be a helpful first step.

**Other Findings and Recommendations:**

- 1) NHGRI should discuss how best to expand awareness of the FAIR Principles and ensure adherence with researchers themselves.
  - o Journal editors stressed that FAIR genomic data sharing is critical to ensuring rigor and reproducibility, which all stakeholders should be keen to preserve and uphold.
- 2) Repositories should be professionally managed, with long-term funding.
  - o Repositories should publish high-level information (metadata) previews for unpublished data sooner.
  - o It would be helpful for repositories to assign accession numbers sooner.
  - o Data license and clear citation information on each record page would be helpful to Journals.
- 3) GitHub alone is not sufficient to meet Journal's needs for dependable sharing of code. Zenodo and Code Ocean are more enduring alternatives, but multiple Journal editors suggested an NIH-backed resource for sharing code.
- 4) Lists of recommended repositories are helpful to researchers and journals:
  - o NIH's "Supplement Information to the NIH Policy for Data Management and Sharing: [Selecting a Repository for Data Resulting from NIH-Supported Research](#)" is a new resource that journals may find helpful to share with researchers.
  - o PLOS One maintains a list of [recommended repositories](#).
  - o GA4GH expressed interest in promoting true and enduring global genomics resources.
- 5) NIH/NHGRI should consider what would happen to genomic data sharing if DNA sequences are reclassified by U.S. regulatory statutes/policies as "identifiable." Europe is working through this issue, and the U.S. may face this in the future.
- 6) Journals are concerned about how the international differences raised in Finding 5 impacts their ability to be fair to all authors, irrespective of where they live and work, as they are having to consider different expectations for authors in different countries.



## Attendee List

<b>NHGRI/NIH Representatives</b>	
<b>Attendee</b>	<b>Affiliation</b>
Allison McCague	NHGRI
Amanda Conti	NHGRI
Ana Stevens	NHGRI
Barbara Thomas	NHGRI
Carolyn Hutter	NHGRI
Chris Gunter	NHGRI
Cristina Kapustij	NHGRI
Dave Kaufman	NHGRI
Elena Ghanaim	NHGRI
Emma Alme	NHGRI
Eric Green	NHGRI
Erin Ramos	NHGRI
Heidi Sofia	NHGRI
Hilary Leeds	OSP/NIH
Jamil B. Scott	NHGRI
Jennifer Strasburger	NHGRI
Jerry Sheehan	NLM/NIH
Jerryl Somani	NHGRI
Julia Slutsman	OER/NIH
Kathryn Funk	NLM/NIH
Ken Wiley	NHGRI
Larry Brody	NHGRI
Lisa Chadwick	NHGRI
Mike Smith	NHGRI
Natalie Kucher	NHGRI
Sara Hull	NHGRI
Stephanie Morris	NHGRI
Steve Sherry	NCBI/NIH
Susan Gregurick	ODSS/NIH
Susan Vasquez	NHGRI
Taunton Paine	OSP/NIH
Valentina Di Francesco	NHGRI
William Maye	NHGRI

<b>Journal Editors/Outside Representatives</b>	
<b>Attendee</b>	<b>Affiliation</b>
Andrew Hufton	<i>Scientific Data</i>

Anthony Long	<i>GENETICS</i>
Barbara Cheifet	<i>Genome Biology</i>
Ben Solomon	<i>American Journal of Medical Genetics/NHGRI</i>
Brooke LaFlamme	<i>Communications Biology</i>
Bruce Korf	<i>American Journal of Human Genetics</i>
Carolina Perdigoto	<i>Nature Communications</i>
Catherine Potenski	<i>Nature Genetics</i>
Bette Phimister	<i>New England Journal of Medicine</i>
Garry Cutting	<i>Human Mutation</i>
Greg Barsh	<i>PLOS Genetics</i>
Gregory Copenhaver	<i>PLOS Genetics</i>
Hillary Sussman	<i>Genome Research</i>
Jan Higgins	<i>Genetics in Medicine</i>
Joerg Heber	<i>PLOS ONE</i>
Lara Szewczak	<i>Cell</i>
Laura Zahn	<i>Science</i>
Magdalena Skipper	<i>Nature</i>
Mark Johnston	<i>GENETICS</i>
Melanie Bahlo	<i>Communications Biology</i>
Michelle Trenkmann	<i>Nature</i>
Mike Bamshad	<i>HGG Advances</i>
Nathalie Le Bot	<i>Nature Communications</i>
Orli Bahcall	<i>Cell Genomics</i>
Quincey Justman	<i>Cell Systems</i>
Rabia Begum	<i>Genome Medicine</i>
Rebecca Furlong	<i>Nature Communications</i>
Richard Sever	<i>bioRxiv</i>
Robert Steiner	<i>Genetics in Medicine</i>
Sara Cullinan	<i>American Journal of Human Genetics</i>
Sarah Callaghan	<i>Patterns</i>
Veronique Kiermer	<i>PLOS</i>
Angela Page	GA4GH
Peter Goodhand	GA4GH

**Appendix 1. Journal Data Policies.**

<i>Communications Biology</i>	<a href="https://www.nature.com/commsbio/journal-policies/editorial-publishing-policies#availability">https://www.nature.com/commsbio/journal-policies/editorial-publishing-policies#availability</a>
<i>GENETICS</i>	<a href="https://www.genetics.org/content/scope-and-publication-policies#data-reagent">https://www.genetics.org/content/scope-and-publication-policies#data-reagent</a>
<i>Genetics in Medicine</i>	<a href="https://www.nature.com/gim/authors-and-referees/copyright-reporting-guidelines-data-policy#research-data-policy">https://www.nature.com/gim/authors-and-referees/copyright-reporting-guidelines-data-policy#research-data-policy</a>  <a href="https://www.nature.com/gim/authors-and-referees/preparation-of-submissions#standardized-nomenclature-of-DNA-sequence-variants-and-variant-data-sharing">https://www.nature.com/gim/authors-and-referees/preparation-of-submissions#standardized-nomenclature-of-DNA-sequence-variants-and-variant-data-sharing</a>
BMC series of journals (e.g., <i>Genome Biology</i> )	<a href="https://www.biomedcentral.com/getpublished/editorial-policies#availability+of+data+and+materials">https://www.biomedcentral.com/getpublished/editorial-policies#availability+of+data+and+materials</a>  <a href="https://genomebiology.biomedcentral.com/articles/10.1186/s13059-020-02010-6">https://genomebiology.biomedcentral.com/articles/10.1186/s13059-020-02010-6</a>
<i>Genome Research</i>	<a href="https://genome.cshlp.org/site/misc/matdatarel.xhtml">https://genome.cshlp.org/site/misc/matdatarel.xhtml</a>
<i>PLOS Genetics</i>	<a href="https://journals.plos.org/plosgenetics/s/data-availability">https://journals.plos.org/plosgenetics/s/data-availability</a>
Nature Portfolio	<a href="https://www.nature.com/nature-research/editorial-policies/reporting-standards#availability-of-data">https://www.nature.com/nature-research/editorial-policies/reporting-standards#availability-of-data</a>