

Ten Recommended Practices for Managing Preprints in Generalist and Institutional Repositories



COAR-ASAPbio Working Group on Preprints in Repositories

December 12, 2022

Acknowledgements: This report was prepared by the [COAR-ASAPbio Working Group on Preprints in Repositories](#) and reviewed by Richard Sever, bioRxiv.

INTRODUCTION

Preprint sharing has become an increasingly common practice, one which has greatly expanded since the COVID-19 pandemic. Posting a preprint makes an article quickly and freely available to everyone (while the published version may take months and be behind a paywall) and opens the manuscript up for transparent community review. The move towards greater preprint sharing is being accompanied by other innovations that leverage the availability of these manuscripts such as [open peer review](#) and [endorsement](#) services, and the ‘publish, review, and curate’ [approach](#) - whereby some [publishers](#) will only review manuscripts already published as preprints. Some [funders](#) are now accepting open peer review on preprints as an alternative to traditional journals. These trends have the potential to drastically improve research communication, making it more efficient, open, and transparent. The [COAR Notify Initiative](#), which is developing a standard protocol for interactions across distributed platforms and services, will support greater linking across services and enable these activities to scale.

There are an increasing number of [dedicated preprint servers](#) (preprint-specific repositories) that are developing new practices to support the unique needs related to managing preprints. However, there are still large gaps in geographic and domain coverage and some authors will choose to deposit their research outputs into another type of repository, such as an institutional or generalist repository. For example, many existing preprint servers only allow submissions in English, which limits the options for researchers who work in other languages. Or, researchers may want to have their preprint hosted in their own country or region to comply with national policies or other preferences. In addition, there is an open question about the [long term funding for many preprint servers](#), as they often rely on short term grant funding, so a distributed network that can support preprint sharing across a range of platforms will contribute to the sustainability of the system. As such, it is clear that institutional and generalist repositories have an important role to play in supporting preprint sharing worldwide.

In 2021, COAR and ASAPbio undertook a survey of current practices and plans regarding the collection of preprints in institutional and generalists repositories (IRs). The [results](#) revealed that approximately two thirds of the responding repositories currently host preprints. However, it was also found that many repositories did not yet support the unique functionalities and practices that are being adopted by preprint servers. To address these gaps, COAR and ASAPbio convened a Working Group to develop recommended practices for

managing preprints¹. The Working Group undertook an environmental scan of existing preprint server functionalities and identified priority practices that should be considered for adoption by any type of repository that is managing preprints.

The Working Group has identified ten recommended practices for managing preprints in generalist and institutional repositories across three areas: linking, discovery, and editorial processes. The practices are listed in the table below, followed by more detailed explanations and concrete examples of how they are currently being implemented. While we acknowledge that many of these practices are not currently in use by institutional and generalist repositories, we hope that these recommendations will encourage repositories around the world that collect preprints to begin to apply them locally.

SUMMARY OF RECOMMENDATIONS

Linking preprint versions, journal versions, and peer reviews

1. Offer a step in the submission process for authors to provide information about other preprint versions, accepted manuscripts, journal versions, and external peer reviews
2. When this information is available, indicate that there is related content in the repository metadata record using “dc:relation” field or “isIdenticalTo”, “isVersionOf”, “isPreprintOf” or “hasReview” and include the PID of the external resource
3. When the information is available, link to related versions and external peer reviews on the landing page of the preprint
4. For each new version of a preprint in the repository, assign a unique PID and include a version number that represents their updates sequentially
5. In addition to versions and reviews, enable authors to link to other related resources such as data, code or other associated outputs

Discoverability

6. Ensure preprints are integrated into domain and preprint discovery and indexing systems
7. Clearly indicate in the metadata record and on the landing page that the document is a preprint, working paper, or other domain specific term

Editorial processes

8. Include a text banner on the landing page of the preprint that informs readers that the document is a preprint
9. Clearly indicate on the landing page or on the repository website what type of moderation or screening processes has been applied to the preprints
10. If a preprint has been removed, retain the metadata and a landing page (tombstone page) that designates its status as “withdrawn”

¹For more general good practices in repositories, please see the [COAR Community Framework for Good Practices in Repositories](#)

DEFINITIONS

Landing page: a human-readable webpage which normally presents most of the metadata in the metadata record and links to the resource (in this case the preprint)

PID (Persistent identifier): a long-lasting reference to a document, file, web page, or other digital object. Most PIDs have a unique identifier which is linked to the current address of the metadata or content

Preprint: A version of a scholarly or scientific paper that precedes peer review

Preprint server: A repository that collects only preprints

Repository: A system that collects and provides access to digital content

Repository record: The machine-readable metadata record

DISCUSSION AND EXAMPLES

Linking preprint versions, journal versions, and peer reviews

Recommendations

1. Offer a step in the submission process for authors to provide information about other preprint versions, accepted manuscripts, journal versions, and external peer reviews
2. When this information is available, indicate that there is related content in the repository metadata record using “dc:relation” field or “isIdenticalTo”, “isVersionOf”, “isPreprintOf” or “hasReview” and include the PID of the external resource
3. When the information is available, link to related versions and external peer reviews on the landing page of the preprint
4. For each new version of a preprint in the repository, assign a unique PID and include a version number that represents their updates sequentially
5. In addition to versions and reviews, enable authors to link to other related resources such as data, code or other associated outputs

Discussion

It is very important for users to know what version of a preprint they are reading, whether it has been published elsewhere, or if there are external peer reviews or endorsements related to that preprint. Wherever possible, these types of relationships should be reflected in the metadata record and the landing page of the preprint.

In order to capture such relationships, we recommend that repositories build into the deposit workflow a step whereby authors can provide information about preprint versions, author's accepted manuscripts, journal versions, peer reviews, and endorsements (as well as other types of related resources such as research data, etc.). This way the author can provide the information directly to the repository, although, in many cases, these resources will only be available after the preprint has been deposited. A number of repositories, such as HAL, the French national repository, attempt to identify related content by searching indexes and discovery services such as CrossRef, Google Scholar, and OpenAIRE and add those links into the metadata record and landing pages. The repository could also provide mechanisms for authors to add this information after the original preprint was deposited, given that many of these resources would be created after the original preprint.

In terms of versioning, a [review of current practices](#) in preprint servers found that practices vary. Some preprint servers, such as arXiv and OSF, assign a DOI to each new version of the preprint. Others, like bioRxiv and medRxiv, maintain all versions under the same DOI, with separate URIs for each version. Zenodo assigns a DOI to each new version and uses a 'concept' DOI to represent all versions. For the purposes of citation, we recommend that a repository assign a unique URI (PID) for each version and include a version number that represents all updates sequentially.

In some cases, two identical copies of a preprint are held in more than one repository. HAL, for instance, in which many French researchers are required to deposit their preprints, maintains copies of preprints that are also held in arXiv and other preprint servers. For the purpose of indexing and user disambiguation, a repository should try to determine if a preprint is an identical copy of one hosted elsewhere, or whether it is unique to the repository. If it is an identical copy, then this should be clearly stated in the repository record and the landing page of the preprint.

It is common practice for a preprint to be submitted to a journal for peer review and publication. Ideally, the repository record and the landing page will also link to an author's accepted manuscript or the journal version of the article. Both bioRxiv and medRxiv, for example, point to the journal versions by searching the titles in Crossref records.

Examples

bioRxiv indicating on the landing page that there are other versions

bioRxiv posts many COVID-19-related papers. A reminder: they have not been formally peer-reviewed and should not guide health-related behavior or be reported in the press as conclusive.

New Results

View current version of this article

Follow this preprint

Within-host evolutionary dynamics and tissue compartmentalization during acute SARS-CoV-2 infection

Mireille Farjo, Katia Koelle, Michael A. Martin, Laura L. Gibson, Kimberly K.O. Walden, Gloria Rendon, Christopher J. Fields, Fadi G. Alnaji, Nicholas Gallagher, Chun Hui Luo, Heba H. Mostafa, Yukari C. Manabe, Andrew Pekosz, Rebecca L. Smith, David D. McManus, Christopher B. Brooke

doi: <https://doi.org/10.1101/2022.06.21.497047>

This article is a preprint and has not been certified by peer review [what does this mean?]

Abstract Full Text **Info/History** Metrics Preview PDF

Abstract

The global evolution of SARS-CoV-2 depends in part upon the evolutionary dynamics within individual hosts with varying immune histories. To characterize the within-host evolution of acute SARS-CoV-2 infection, we deep sequenced saliva and nasal samples

Subject Area

Microbiology

Subject Areas

All Articles

Animal Behavior and Cognition

Biochemistry

Article Information

doi: <https://doi.org/10.1101/2022.06.21.497047>

History June 22, 2022.

Article Versions

You are currently viewing Version 1 of this article (June 22, 2022 - 1431).

medRxiv preprint linking to the published journal article

medRxiv THE PREPRINT SERVER FOR HEALTH SCIENCES

CSH Cold Spring Harbor Laboratory BMJ Yale

HOME | SUBMIT | FAQ | BLOG | ALERTS | RSS | ABOUT

Search

Advanced Search

Previous Next

Posted January 31, 2021.

Download PDF **Print/Save Options** **Email** **Share** **Author Declarations** **Citation Tools** **Data/Code**

Like 0

Surveillance-to-Diagnostic Testing Program for Asymptomatic SARS-CoV-2 Infections on a Large, Urban Campus - Georgia Institute of Technology, Fall 2020

Greg Gibson, Joshua S. Weitz, Michael P. Shannon, Benjamin Holton, Anton Bryksin, Brian Liu, Sandra Bramblett, JulieAnne Williamson, Michael Farrell, Alexander Ortiz, Chaouki T. Abdallah, Andrés J. García

doi: <https://doi.org/10.1101/2021.01.28.21250700>

Now published in Epidemiology doi: 10.1097/ede.0000000000001448

Abstract Full Text Info/History Metrics Preview PDF

Abstract

A SARS-CoV-2 testing program combining pooled saliva sample surveillance leading to diagnosis and intervention surveyed over 112,000 samples from 18,029 students, staff and faculty, as part of integrative efforts to mitigate transmission at the Georgia Institute of Technology in Fall 2020. Cumulatively, 1,508 individuals were confirmed diagnostically. The surveillance strategy, including focused intensification of testing

Subject Area

Infectious Diseases (except HIV/AIDS)

Subject Areas

arXiv record with link to overlay journal, *Advances in Combinatorics*

Cornell University

We gratefully acknowledge support from the Simons Foundation and member institutions.

arXiv > math > arXiv:2107.08658

Search... All fields Search

Help | Advanced Search

Mathematics > Combinatorics

[Submitted on 19 Jul 2021 (v1), last revised 22 Jul 2022 (this version, v3)]

Extremal functions for sparse minors

Kevin Hendrey, Sergey Norin, David R. Wood

The "extremal function" $c(H)$ of a graph H is the supremum of densities of graphs not containing H as a minor, where the "density" of a graph G is the ratio of the number of edges to the number of vertices. Myers and Thomason (2005), Norin, Reed, Thomason and Wood (2020), and Thomason and Wales (2019) determined the asymptotic behaviour of $c(H)$ for all polynomially dense graphs H , as well as almost all graphs H of constant density. We explore the asymptotic behavior of the extremal function in the regime not covered by the above results, where in addition to having constant density the graph H is in a graph class admitting strongly sublinear separators. We establish asymptotically tight bounds in many cases. For example, we prove that for every planar graph H ,

$$c(H) = (1 + o(1)) \cdot \max \left\{ \frac{|V(H)|}{2}, |V(H)| - \alpha(H) \right\},$$

extending recent results of Haslegrave, Kim and Liu (2020). We also show that an asymptotically tight bound on the extremal function of graphs in minor-closed families proposed by Haslegrave, Kim and Liu (2020) is equivalent to a well studied open weakening of Hadwiger's conjecture.

Subjects: **Combinatorics (math.CO)**
Cite as: **arXiv:2107.08658 [math.CO]**
(or **arXiv:2107.08658v3 [math.CO]** for this version)
<https://doi.org/10.48550/arXiv.2107.08658>

Journal reference: **Advances in Combinatorics**, 2022 5, 43pp
Related DOI: <https://doi.org/10.19086/aic.2022.5>

Download:

- PDF
- Other formats (license)

Current browse context: **math.CO**

< prev | next >
new | recent | 2107

Change to browse by: **math**

References & Citations

- NASA ADS
- Google Scholar
- Semantic Scholar

Export BibTeX Citation

Bookmark

HAL record with link to identical arXiv version

HAL

New Interface

en Sign in

hal-00000820, version 1

Preprints, Working Papers, ...

Hodge-DeRham theory with degenerating coefficients

Fouad Elzein¹ Details

¹ LMJL - Laboratoire de Mathématiques Jean Leray

Abstract : Let $\mathcal{L}$ be a local system on the complement X^* of a normal crossing divisor (NCD) Y in a smooth analytic variety X and let $j : X^* = X - Y \rightarrow X$ denotes the open embedding. The purpose of this paper is to describe a weight filtration W on the direct image $j_*\mathcal{L}$ and in case a morphism $f : X \rightarrow D$ to a complex disc is given with $Y = f^{-1}(0)$, the weight filtration on the complex of nearby cocycles $\Psi_f(\mathcal{L})$ on Y . A comparison theorem shows that the filtration coincides with the weight defined by the logarithm of the monodromy and provides the link with various results on the subject.

Mots-clés : Hodge mixte faisceau pervers décomposition monodromie pureté décomposition.

Document type : Preprints, Working Papers, ...

Domain : Mathematics [math] / Algebraic Geometry [math.AG]

Complete list of metadata

Cited literature [10 references]

FILES

Hodge.pdf

IDENTIFIERS

- HAL id : hal-00000820, version 1
- ARXIV : math.AG/0311083

COLLECTIONS

UNIV-NANTES | LMJL | CNRS | FMPL | INSMI | NANTES-UNIVERSITE

Discovery

Recommendations

6. Ensure preprints are integrated into domain and preprint discovery and indexing systems
7. Clearly indicate in the metadata record and on the landing page that the document is a preprint, working paper, or other domain specific term

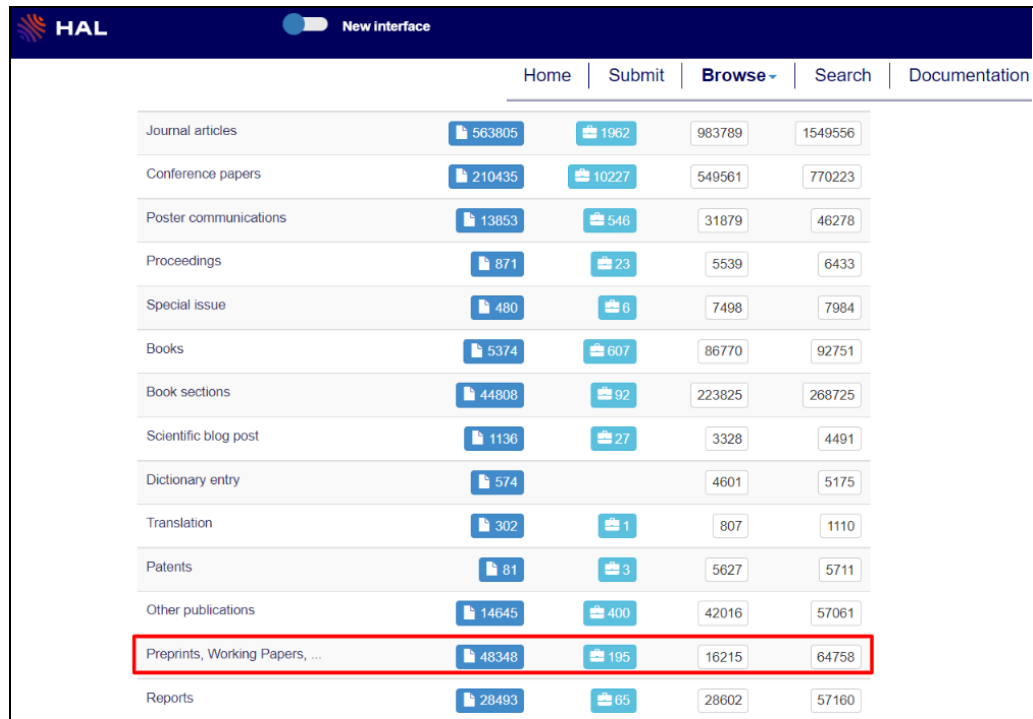
Discussion

Discovery of preprints is critical as many authors post preprints in order to share their work with their peers. To the extent possible, preprints in the repository should be integrated into relevant domain discovery systems. Repositories hosting original preprints should seek to have their preprints indexed by relevant search platforms, such as Europe PMC, which aggregates preprints in the biomedical fields. Both [PubMed Central](#) and [Europe PMC](#) have specified eligibility considerations for indexing preprints. Institutional and generalist repositories, which contain a variety of content types will need to direct these discovery systems to the relevant records only, which can be done by creating a “preprint set” in the OAI-PMH record. Many domain preprint services are minting CrossRef DOIs for preprints, enabling them to be more discoverable as they are then searchable in the CrossRef metadata.

Unambiguously labeling an article as a “preprint” (or other domain specific term such as “working paper”) will ensure readers and discovery systems are aware of the status of the article and allows repositories to implement any preprint-specific features related to that item. Repositories should enable authors to tag their articles as preprints and include this information in the metadata record. This can be facilitated by integrating the [COAR Resource Type](#) controlled vocabulary into the repository platform, which standardizes the terms used and contributes to greater consistency in the repository and across the ecosystem. The COAR Resource Type vocabulary contains preferred and alternative labels for preprints and other domain specific terms, which have been translated into several languages.

Examples

HAL listing publication types in the repository



HAL New interface

Home | Submit | **Browse** | Search | Documentation

Journal articles	563805	1962	983789	1549556
Conference papers	210435	10227	549561	770223
Poster communications	13853	546	31879	46278
Proceedings	871	23	5539	6433
Special issue	480	6	7498	7984
Books	5374	607	86770	92751
Book sections	44808	92	223825	268725
Scientific blog post	1136	27	3328	4491
Dictionary entry	574		4601	5175
Translation	302	1	807	1110
Patents	81	3	5627	5711
Other publications	14645	400	42016	57061
Preprints, Working Papers, ...	48348	195	16215	64758
Reports	28493	65	28602	57160

Editorial Processes

Recommendations

8. Include a text banner on the landing page of the preprint that informs readers that the document is a preprint.
9. Clearly indicate on the landing page or on the repository website what type of moderation or screening processes have been undertaken by the repository
10. If a preprint has been removed, retain the metadata and a landing page (tombstone page) that designates its status as “withdrawn”

Discussion

It is important for readers to be aware of the status of a preprint and what level of screening has been carried out by the repository. The need for screening will vary greatly depending on the field. In the biomedical sciences, there is a more obvious need to have clear disclaimers that caution readers (see ASAPbio’s [Guiding principles for preprint servers on preprint labeling](#)) while in other domains (physics, math) this may not be necessary. In addition, there is increasing ambiguity around whether a preprint has undergone peer review, as there are a growing number of external peer-review services and overlay journals (e.g. [Peer Community In](#), [Review Commons](#), [Episciences](#)). In this case, the preprint may have been peer reviewed, and may be published in a journal with little or no modifications. While it is common for institutional and generalist repositories to apply lightweight curatorial processes to check for metadata completeness, they rarely review the scientific integrity of the resources they host. Therefore we recommend that repositories include a banner indicating that the record may not have been peer reviewed. We suggest something like, **“This article has been labeled as a preprint and may not have been peer-reviewed”**. In addition, to ensure transparency around the publication process, it is also recommended that the repository provides information on its website about what type of screening (or lack thereof) is undertaken prior to posting and this information could also be linked to from the banner on the individual landing pages.

Preprints are part of the permanent research record, however, if at a later date, authors no longer stand by the scientific claims contained in a paper, it should be possible to remove it from the repository and label the preprint record as withdrawn. In this case, the repository record should remain and associated metadata should be updated to indicate the status as withdrawn. In addition, repositories should have processes in place to address situations where a posted preprint breaches the policies of the repository, represents a legal challenge, or could incur substantial risk of harm (e.g. to society, specific groups or an individual). In such rare circumstances, the full-text preprint should be removed from the repository, with an indication in the metadata record and the landing page (tombstone record) that the preprint has been removed. In addition, the repository should have a public policy related to the withdrawal and removal of resources.

Examples

medRxiv landing page disclaimer about the preprint



medRxiv THE PREPRINT SERVER FOR HEALTH SCIENCES

CSH Cold Spring Harbor Laboratory BMJ Yale

HOME | SUBMIT | FAQ | BLOG | ALERTS / RSS | ABOUT

Search

Advanced

Immunogenicity and safety of a 4th homologous booster dose of a SARS-CoV-2 recombinant spike protein vaccine (NVX-CoV2373): a phase 2, randomized, placebo-controlled trial

Katia Alves, Joyce S Pleased, Shirley Galbiati, Gordon Chau, Shane Cloney-Clark, Mingzhu Zhu, Raj Kalkeri, Nita Patel, Kathy Smith, Alex Marcheschi, Susan Pfeiffer, Heather McFall, Gale Smith, Gregory M. Glenn, Filip Dubovsky, Raburn M. Mallory the Novavax 2019nCoV-101 Study Group

doi: <https://doi.org/10.1101/2022.11.18.22282414>

This article is a preprint and has not been peer-reviewed [what does this mean?]. It reports new medical research that has yet to be evaluated and so should not be used to guide clinical practice.

Previous

Posted November 20, 2022.

Download PDF

Print/Save Options

Author Declarations

Supplementary Material

Data/Code

Email

Share

Citation Tools

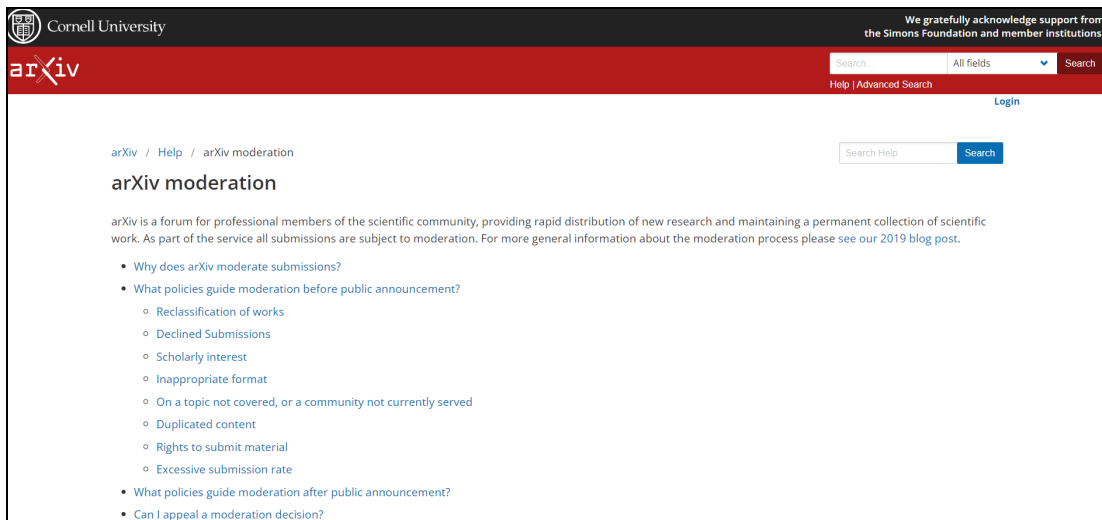
Like 13

COVID-19 SARS-CoV-2 preprints from medRxiv and bioRxiv

Subject Area

Abstract Full Text Info/History Metrics Preview PDF

arXiv moderation process



Cornell University

We gratefully acknowledge support from the Simons Foundation and member institutions.

arXiv

Search All fields Search

Help | Advanced Search

Login

arXiv / Help / arXiv moderation

arXiv moderation

arXiv is a forum for professional members of the scientific community, providing rapid distribution of new research and maintaining a permanent collection of scientific work. As part of the service all submissions are subject to moderation. For more general information about the moderation process please see our 2019 blog post.

- Why does arXiv moderate submissions?
- What policies guide moderation before public announcement?
 - Reclassification of works
 - Declined Submissions
 - Scholarly interest
 - Inappropriate format
 - On a topic not covered, or a community not currently served
 - Duplicated content
 - Rights to submit material
 - Excessive submission rate
- What policies guide moderation after public announcement?
- Can I appeal a moderation decision?

Withdrawn preprint from [bioRxiv](#)

bioRxiv
THE PREPRINT SERVER FOR BIOLOGY

[HOME](#) | [SUBMIT](#) | [FAQ](#) | [BLOG](#) | [ALERTS / RSS](#) | [ABOUT](#) | [CHANNELS](#)

Advanced Search

[Previous](#)
[Next](#)

Posted August 28, 2019.

[Download PDF](#)
[Email](#)

[Print/Save Options](#)
[Share](#)

[Citation Tools](#)

[Like 0](#)

COVID-19 SARS-CoV-2 preprints from medRxiv and bioRxiv

Subject Area

Genomics

Subject Areas

All Articles

Animal Behavior and Cognition

Biochemistry

bioRxiv posts many COVID19-related papers. A reminder: they have not been formally peer-reviewed and should not guide health-related behavior or be reported in the press as conclusive.

WITHDRAWN [Follow this preprint](#)

Rewiring of the cellular and inter-cellular landscape of the human colon during ulcerative colitis

Christopher S. Smillie, Moshe Biton, Jose Ordovas-Montanes, Keri M. Sullivan, Grace Burgin, Daniel B. Graham, Rebecca H. Herbst, Noga Rogel, Michal Slyper, Julia Waldman, Malika Sud, Elizabeth Andrews, Adam L. Haber, Sanja Vickovic, Danielle Dionne, Lan T. Nguyen, Alexandra Chloé Villani, Matan Hofree, Elizabeth A. Creasey, Hailiang Huang, Orit Rozenblatt-Rosen, John J. Garber, Hamed Khalili, A. Nicole Desch, Mark J. Daly, Ashwin N. Ananthakrishnan, Alex K. Shalek, Ramnik J. Xavier, Aviv Regev

doi: <https://doi.org/10.1101/455451>

This article is a preprint and has not been certified by peer review [what does this mean?].

[1](#)
[0](#)
[0](#)
[0](#)
[5](#)
[0](#)
[37](#)

Abstract Info/History Metrics [Preview PDF](#)

Abstract

The paper has been withdrawn owing to erroneous inclusion of confidential information relating to a third party.