Enabling preprint discovery, evaluation, and analysis with Europe PMC

Mariia Levchenko[1]*, Michael Parkin[1], Johanna McEntyre[1], Melissa Harrison[1]

[1]European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Wellcome Genome Campus, Hinxton, United Kingdom

* Corresponding author
E-mail: levchmar@ebi.ac.uk (ML)

Abstract

Preprints provide an indispensable tool for rapid and open communication of early research findings. Preprints can also be revised and improved based on scientific commentary uncoupled from journal-organised peer review. The uptake of preprints in the life sciences has increased significantly in recent years, especially during the COVID-19 pandemic, when immediate access to research findings became crucial to address the global health emergency.

With ongoing expansion of new preprint servers, improving discoverability of preprints is a necessary step to facilitate wider sharing of the science reported in preprints. To address the challenges of preprint visibility and reuse, Europe PMC, an open database of life science literature, began indexing preprint abstracts and metadata from several platforms in July 2018. Since then, Europe PMC has continued to increase coverage through addition of new servers, and expanded its preprint initiative to include the full text of preprints related to COVID-19 in July 2020 and then the full text of preprints supported by the Europe PMC funder consortium in April 2022. The preprint collection can be searched via the website and programmatically, with abstracts and the open access full text of COVID-19 and Europe PMC funder preprint subsets available for bulk download in a standard machine-readable JATS XML format. This enables automated information extraction for large-scale analyses of the preprint corpus, accelerating scientific research of the preprint literature itself.

This publication describes steps taken to build trust, improve discoverability, and support reuse of life science preprints in Europe PMC. Here we discuss the benefits of indexing preprints alongside peer-reviewed publications, and challenges associated with this process.

Introduction

Preprint adoption in the life sciences

Traditionally, scientific results are communicated by publishing a peer-reviewed journal article. However, the current publishing process has a number of challenges and limitations. Increasing costs to publish open access, and to access the published literature, present a problem for both authors and readers [1]. The publishing process itself can be quite lengthy. It can take several months, and in some cases years, from submission to publication [2]. If the initial submission is rejected, additional time and effort is needed to re-submit to another journal. Editorial and peer review decisions can be subjective [3] and subject to gender, reputational, and diversity biases [4-7]. If mistakes, or even fraud, take place they are often hard to correct. Studies suggest that many publications are not retracted when they should be, despite an increase in publication retractions in recent years [8-10].

Preprints can provide a way to address some of these limitations. Preprints are complete scientific manuscripts without journal-organised peer review, uploaded by the authors to a public server [11]. Preprints offer several advantages, such as rapid communication and no charge to read and post. Preprints also provide an opportunity for early feedback and additional exposure prior to final publication in a journal [12-13].

In some disciplines, such as physics, preprints have been part of research communication for decades [14]. Efforts to promote preprints in biology trace back several decades [15-16]. Nonetheless, in the life sciences, it is still a relatively new path to share research findings. Preprints represented less than 1% of the total number of journal articles in the life sciences published in 2016 (Fig 1). Emergence of life science-specific preprint servers supported preprint adoption in the community [17-19]. Changes to publisher and funder policies have further incentivised preprinting [19]. Since 2016 several research funding organisations have opted to take preprints into consideration in funding decisions, and currently encourage or

require investigators to post preprints prior to publication [19]. A 2020 study on peer review and preprint policies across a subset of major journals found that in the life and earth sciences 91% of surveyed journals would accept preprinted submissions [20], and many journals actively encourage it (https://asapbio.org/journal-policies). Several journals make manuscripts available to read before journal-organised peer review [17, 21-23]. Such publications can therefore also be considered preprints. More recently, the 'Publish, Review, Curate' (PRC) peer-review model has been pioneered by eLife and later adopted by GigaByte Journal [24-25]. In this model every manuscript sent for review is published as a "Reviewed Preprint" and is available alongside an editorial assessment, public reviews, and a response from the authors. Other innovative publisher practices that incentivise preprinting include exclusively reviewing preprints, extended scoop protection or publication charge discounts (https://asapbio.org/journal-policies). Finally, preprint adoption by life science researchers has been heavily supported through advocacy work. Organisations such as ASAPbio have promoted the use of preprints for research dissemination.

As preprints became more established in the life sciences their growth rate increased significantly. Between 2016 and 2018, the number of life science preprints was growing 10 times faster than the number of journal articles [19, 26]. In 2023 the share of life science preprints had grown to over 12% (Fig 1).

Fig 1. Percentage of preprints in Europe PMC posted yearly between 2013–2023. The number of preprints is divided by the total number of journal articles in PubMed published in the same period. Note that the numbers for preprints are based on the content from servers indexed in Europe PMC. Accessed 8 April 2024.

Growth of preprints during the COVID-19 pandemic

Preprints enable swift dissemination of scientific findings. Therefore, their importance becomes increasingly apparent in public health crises. Some funders and institutions, including the Bill & Melinda Gates Foundation, Médecins Sans Frontières, the US National

Institute of Health, and Wellcome, have explicitly encouraged preprinting during outbreaks caused by Zika and Ebola viruses. However, preprint adoption during these epidemics remained relatively limited [27].

The COVID-19 pandemic posed a truly unprecedented challenge for scholarly communications. It required rapid sharing of research results to inform policy and clinical decision making. In 2020 stakeholders, such as the World Health Organization (WHO), Organisation for Economic Co-operation and Development (OECD), national science and technology advisors, and funding institutions, encouraged immediate open sharing of COVID-19 related publications and data [28-29]. Preprints played an important role in early dissemination of COVID-19 science. Within the first few months of the pandemic they comprised over 30% of the published COVID-19 research (Fig 2). This was largely due to the higher speed of sharing preprints compared to journal publications despite many in the scholarly publishing community supporting this paradigm. Publisher's efforts included implementation of an expedited peer review and publication process, waiver of author processing charges (APCs), and open access to publications on COVID-19 [30]. In March 2020 more than 50 publishers made COVID-19 full text articles available via PubMed Central (PMC), and through PMC International, via Europe PMC (https://www.ncbi.nlm.nih.gov/labs/pmc/about/covid-19/). Several funding agencies also supported preprint adoption during the COVID-19 pandemic, encouraging or requiring research results to be shared as preprints prior to publication (https://asapbio.org/funder-policies).

Fig 2. Preprints comprise a significant portion of COVID-19 research. Percentage of COVID-19 preprints (in blue) and COVID-19 journal articles (in green) available in Europe PMC by month. This data includes preprints that have been subsequently published as a journal article. Accessed 8 April 2024.

Facilitating preprint discovery with Europe PMC

Despite their significance, preprints can be harder to find and reuse than journal articles. Preprints are typically scattered across many platforms. Over 60 preprint servers have a biomedical or medical scope (https://asapbio.org/preprint-servers). Preprint coverage across different scholarly discovery engines greatly varies. One dataset reports coverage under 80% surveyed across various tools for at least 30% of preprint servers (https://tinyurl.com/searchenginecomparison). Many of the discovery platforms are proprietary. Such platforms may lack community governance and sustainability plans for long term availability. This has been exemplified by the discontinuation of Microsoft Academic in 2021 and Meta in 2022 [31-32]. Lack of transparency in indexing policies and algorithms may also pose a problem. For example, reproducibility in search strategies is key when using such tools for systematic reviews [33]. Several discovery tools offer free options for preprint searches to individual researchers. However, few of them provide free programmatic access and bulk download options. This can pose difficulties for text-mining research involving preprints.

Recognising the importance of preprints in the life sciences and the existing challenges of preprint discovery and reuse, in 2018 Europe PMC started indexing life science preprints alongside peer-reviewed publications [34]. Europe PMC (https://europepmc.org/) is an open science platform and a life science literature database [35]. It is supported by a group of 35 international science funders as their repository of choice. Europe PMC is developed by the European Bioinformatics Institute (EMBL-EBI). It is part of the PubMed Central (PMC) International archive network, built in collaboration with the PMC archive in the USA. Europe PMC provides access to millions of abstracts and full text records from a range of sources, including MEDLINE (PubMed), PMC, and Agricola, as well as many preprint servers. All open access content is freely available via the Europe PMC website, Application Programming Interfaces (APIs) and as bulk downloads.

In July 2020 Europe PMC enhanced its preprint coverage by indexing the full text of preprints related to COVID-19. The full text was made available in a standard JATS XML format to support text-mining analyses of the COVID-19 literature [36-37]. This corpus can be accessed via the Europe PMC website, APIs, and bulk downloads, where licensing allows. This initiative aimed to increase the discoverability of COVID-19-related research and concluded in October 2023. It complemented other projects providing access to COVID-19-related literature [38-39].

From April 2022 the full text of preprints that acknowledge funding from one or more of the Europe PMC funders from selected servers was also made available in Europe PMC [40]. This initiative aimed to expand the collection of full text preprints for future analyses. Indexing the full text of such preprints also improves visibility of science supported by Europe PMC funders. In a similar effort in January 2023, NIH-affiliated and supported preprints from selected servers became available via PMC USA and PubMed [41].

This paper describes the scope and technical approach to indexing life science preprints in Europe PMC. It summarises our efforts to improve preprint discoverability and help build trust in preprints. We also share lessons learnt in the process.

Methods

Technical implementation

Preprint selection criteria

Europe PMC includes abstracts and metadata for preprints from preprint servers that fulfil the following criteria (https://europepmc.org/Preprints#preprint-criteria).

- **Scope:** the preprint server should have a significant proportion of life science or interdisciplinary subjects preprints.

- **Volume:** the preprint server should have at least 30 preprints available.
- **Screening**: the preprint server should have a screening procedure detailed in a public statement.
- **Access to metadata:** preprint metadata should be available in machine-readable format, e.g. via an API.
- **Access to full text:** the full text of all preprints should be visible on the preprint server with no barrier to access.
- **Publication ethics**: the preprint server should have a public statement on plagiarism, misconduct, and competing interests policies.

In addition, Europe PMC requires the following metadata in a machine readable format:

- Preprint identifier (currently a Crossref DOI is required)
- Preprint title
- Author names
- Abstract
- Publication date

As of 4 April 2024, Europe PMC indexes preprints from 32 preprint servers (Table 1).

Table 1. The list of preprint servers indexed in Europe PMC.

Preprint server/platform name	Search syntax in Europe PMC	Preprint server or platform URL	Preprint metadata and abstracts indexed	Preprint full text indexed*
Access Microbiology	PUBLISHER:"Access Microbiology"	https://www.microbiologyresearch.or	All preprints	No

		g/content/journal/acmi		
AfricArXiv	PUBLISHER:"AfricArXiv"	https://osf.io/preprints/africarxiv/	All preprints	No
agriRxiv	PUBLISHER:"agriRxiv"	https://agrirxiv.org/	All preprints	No
AIJR Preprints	PUBLISHER:"AIJR Preprints"	https://preprints.aijr.org/	All preprints	No
ARPHA Preprints	PUBLISHER:"ARPHA Preprints"	https://preprints.arphahub.com/	All preprints	No
arXiv	PUBLISHER:"arXiv"	https://arxiv.org/	COVID-19 preprints	COVID-19 preprints
Authorea Preprints	PUBLISHER:"Authorea Preprints"	https://authorea.com/	All preprints	COVID-19 preprints
Beilstein Archives	PUBLISHER:"Beilstein Archives"	https://www.beilstein-archives.org/xiv/home	All preprints	No
BioHackrXiv	PUBLISHER:"BioHackrXiv"	https://biohackrxiv.org/	All preprints	No
bioRxiv	PUBLISHER:"bioRxiv"	https://www.biorxiv.org/	All preprints	COVID-19 and CC-licensed

				Europe PMC funder preprints
ChemRxiv	PUBLISHER:"ChemRxiv"	https://chemrxiv.org/	All preprints	COVID-19 preprints
EcoEvoRxiv	PUBLISHER:"EcoEvoRxiv"	https://ecoevorxiv.org/	All preprints	No
Emerald Open Research	PUBLISHER:"Emerald Open Res"	https://emeraldopenresearch.com/	All preprints	No
F1000 Research	PUBLISHER:"F1000Res"	https://f1000research.com/	All preprints	No
Gates Open Research	PUBLISHER:"Gates Open Res"	https://gatesopenresearch.org/	All preprints	No
Health Open Research	PUBLISHER:"AMRC Open Res" OR PUBLISHER:"healthopenres"	https://healthopenresearch.org/	All preprints	No
HRB Open Research	PUBLISHER:"HRB Open Res"	https://hrbopenresearch.org/	All preprints	No
MedEdPublish	PUBLISHER:"MedEdPublish"	https://www.mededpublish.org/	All preprints	No
medRxiv	PUBLISHER:"medRxiv	https://www.medrx	All preprints	COVID-19

	"	iv.org/		and CC-licensed Europe PMC funder preprints
NIHR Open Research	PUBLISHER:"NIHR Open Res"	https://openresearch.nihr.ac.uk/	All preprints	No
Open Research Africa	PUBLISHER:"AAS Open Res" OR PUBLISHER:"Open Res Africa"	https://openresearchafrica.org/	All preprints	No
Open Research Europe	PUBLISHER:"Open Res Europe"	https://open-research-europe.ec.europa.eu/	All preprints	No
PaleorXiv	PUBLISHER:"PaleorXiv"	https://osf.io/preprints/paleorxiv/	All preprints	No
PeerJ Preprints	PUBLISHER:"PeerJ Preprints"	https://peerj.com/preprints/	All preprints	No
Preprints.org	PUBLISHER:"Preprints.org"	https://www.preprints.org/	All preprints	COVID-19 preprints
PsyArXiv	PUBLISHER:"PsyArXiv"	https://psyarxiv.com/	All preprints	COVID-19 preprints

Qeios	PUBLISHER:"Qeios"	https://www.qeios.com/	All preprints	No
Research Square	PUBLISHER:"Research Square"	https://www.researchsquare.com	All preprints	COVID-19 and Europe PMC funder preprints
SciELO Preprints	PUBLISHER:"SciELO Preprints"	https://preprints.scielo.org/	All preprints	COVID-19 preprints
ScienceOpen Preprints	PUBLISHER:"Science Open"	https://www.scienceopen.com/collection/5916e67c-0edf-472a-ad8e-6e205a4e080d	All preprints	No
SSRN	PUBLISHER:"SSRN"	https://www.ssrn.com/index.cfm/en/	COVID-19 preprints	COVID-19 preprints
Wellcome Open Research	PUBLISHER:"Wellcome Open Res"	https://wellcomeopenresearch.org/	All preprints	No

Footnote: *COVID-19 preprints posted until 1 November 2023. Europe PMC funder preprints posted on or after 1 April 2022.

In addition to abstracts and metadata, Europe PMC includes full text of preprints related to COVID-19, as well as preprints supported by the Europe PMC funder consortium.

176 The COVID-19 preprints were selected from a subset of preprint servers indexed in Europe
177 PMC (Table 1). Coronavirus-related preprints were identified using the following metadata
178 search query in Europe PMC: (TITLE_ABS:"COVID" OR TITLE_ABS:"SARS CoV 2" OR
179 TITLE_ABS:"2019-nCoV" OR TITLE_ABS:coronavirus* OR TITLE_ABS:"corona virus" OR
180 TITLE_ABS:"MERS CoV" OR TITLE_ABS:"Middle East Respiratory Syndrome" OR
181 TITLE_ABS:"Severe Acute Respiratory Syndrome"). In the case of arXiv and SSRN servers,
182 a curated list of COVID-19 preprints was provided by the corresponding server instead. In
183 addition to criteria listed above, contact details for the corresponding author were a
184 prerequisite for indexing the preprint full text in Europe PMC. Preprints and any new versions
185 posted before 1 November 2023 were included, thereafter the initiative ended.

186 Europe PMC funder preprints are selected from the medRxiv, bioRxiv, and Research Square
187 servers. Those preprints are identified based on the full text funding acknowledgement. The
188 Acknowledgement or Funding section in the PDF file of the full text preprint is text-mined for
189 Europe PMC funder names using a series of Python scripts. Support from at least one of the
190 Europe PMC funders (https://europepmc.org/Funders/) has to be disclosed. Only preprints
191 that have a Creative Commons licence are selected for full text indexing. Preprints and any
192 new versions posted on or after 1 April 2022 are included.

193 Preprint ingest and conversion process

194 Crossref (https://www.crossref.org/) is a DOI registration agency for scholarly documents,
195 including preprints. Europe PMC uses the Crossref REST API to retrieve new preprint
196 abstracts and metadata as well as updates on a daily basis [42]. Relevant preprint content is
197 identified based on several filters including the DOI prefix and work type. Preprint records
198 added to the Europe PMC database are assigned a unique PPR identifier.

199 To index and display full text for COVID-19 and Europe PMC funder preprints, Europe PMC
200 has set up a processing workflow. The open source Europe PMC plus manuscript
201 submission system is used to process full text preprints [43].

202 In this workflow PDF documents for each preprint version are obtained from selected
203 preprint servers (Table 2). The files include the manuscript, figures, and supplementary
204 information.

205 **Table 2. Methods of PDF/metadata retrieval for full text preprints by server.**

Preprint server	**Method of PDF and metadata retrieval**	
	COVID-19 preprints	**Europe PMC funder preprints**
arXiv	Custom metadata feed + DataCite API	Not applicable
Authorea Preprints	Website	Not applicable
bioRxiv	bioRxiv API + website	bioRxiv API + website
ChemRxiv	ChemRxiv API	Not applicable
medRxiv	medRxiv API + website	medRxiv API + website
Preprints.org	Crossref API + website	Not applicable
PsyArXiv	OSF API	Not applicable
Research Square	Research Square API	Research Square API

SciELO Preprints	Website	Not applicable
SSRN	Custom metadata feed	Not applicable

Using the Europe PMC plus submission system, the preprint full text in PDF format and associated files are passed to an external vendor to create the full text JATS XML. Full text preprints are tagged with the article-type attribute "preprint" to differentiate preprints from journal-published articles. The full text HTML version of the preprint is created after a further quality assurance step by an external vendor to resolve any XML errors.

Full text in JATS XML format for all COVID-19 and Europe PMC funder preprints is indexed for search. Further redistribution is dependent on the permissions of the individual preprint license. This includes display, API access and download options. Preprint license information is identified during metadata retrieval from the preprint server and is captured in the preprint XML during processing. The full text for preprints with a Creative Commons license is automatically released for display and text mining after 14 days. The corresponding author has an option to review the preprint full text during this time. For preprints where no Creative Commons license could be identified, we contact the corresponding author. They are asked to approve the full text, and to add a license to the version of their preprint in Europe PMC that would allow unrestricted reuse for research purposes (https://plus.europepmc.org/user-guide/preprintfaqs#license). If such a license is added, the full text is displayed on the Europe PMC website and made available via the API and bulk downloads as part of the open access subset.

Handling preprint updates

Linking preprints to published journal articles

Preprints in Europe PMC are linked to the corresponding journal-published versions. The link is established for Europe PMC-indexed journal articles that have a PubMed ID (PMID). For some preprint records the DOI for the journal article is provided by the preprint server in the Crossref metadata. For others, an automated process in Europe PMC links preprints to journal articles based on matching titles and first author surnames. The title matching accounts for changes in text case and punctuation.

Handling preprint versions

There are two main ways in which preprint servers indexed in Europe PMC assign persistent identifiers to preprint versions (Table 3):

- Each version of the preprint is assigned its own unique DOI or identifier.
- All versions of the preprint are registered under the same DOI or identifier.

Preprint versions that have a unique DOI or other identifier each receive their own PPRID in Europe PMC. The existence of new versions is detected via Crossref metadata updates or via specific preprint server APIs. In this case previous versions are linked to the most recent version using an internally developed matching process. When a new version is added, only this latest version will be indexed and available in search results. Previous versions can be accessed through links from the latest version. Preprint versions that share a single DOI or other identifier have a single PPRID in Europe PMC, and only the latest record is available in Europe PMC. No links are established to the previous versions and the version history is not available for such records. The original publication date for the version 1 is retained for such records in Europe PMC.

Table 3. Preprint versioning type and version availability in Europe PMC by server.

Preprint server or platform name	Versioning type	Preprint version information in Europe PMC
Access Microbiology	Unique identifier for each version	Available
AfricArXiv	Single identifier for all versions	Not available
agriRxiv	Single identifier for all versions	Not available
AIJR Preprints	Single identifier for all versions	Not available
ARPHA Preprints	Single identifier for all versions	Not available
arXiv.org	Unique identifier for each version	Available
Authorea Preprints	Unique identifier for each version	Available
Beilstein Archives	Unique identifier for each version	Available
BioHackrXiv	Unique identifier for each version	Not available
bioRxiv	Single identifier for all versions	Not available
ChemRxiv	Unique identifier for each version and concept identifier for latest version available	Available

EcoEvoRxiv	Single identifier for all versions	Not available
Emerald Open Research	Unique identifier for each version	Available
F1000 Research	Unique identifier for each version	Available
Gates Open Research	Unique identifier for each version	Available
Health Open Research	Unique identifier for each version	Available
HRB Open Research	Unique identifier for each version	Available
MedEdPublish	Unique identifier for each version	Available
medRxiv	Single identifier for all versions	Not available
NIHR Open Research	Unique identifier for each version	Available
Open Research Africa	Unique identifier for each version	Available
Open Research Europe	Unique identifier for each version	Available
PaleorXiv	Single identifier for all versions	Not available

PeerJ Preprints	Unique identifier for each version and concept identifier for latest version available	Available
Preprints.org	Unique identifier for each version	Available
PsyArXiv	Single identifier for all versions	Not available
Qeios	Unique identifier for each version	Available
Research Square	Unique identifier for each version	Available
SciELO Preprints	Single identifier for all versions	Not available
ScienceOpen Preprints	Unique identifier for each version	Available
SSRN	Single identifier for all versions	Not available
Wellcome Open Research	Unique identifier for each version	Available

Handling preprint withdrawals and removals

Preprints may be withdrawn or removed from the preprint server due to a variety of reasons [44-45]. When that happens, typically a withdrawal or removal notice is added to the preprint record as a new version of the preprint. Such notices are often only a few sentences long. Therefore, in Europe PMC those notices are identified based on document length for the full text preprint subset [43]. Withdrawn and removed preprint records are then flagged,

manually checked, and tagged with an appropriate article-type attribute in the JATS XML. Once the article-type is correctly tagged, the notice full text is released for display and linking to previous versions on Europe PMC.

In some cases the preprint URL leads to a 404 page, and not even a notification remains. If this is reported to the Europe PMC team the record is removed from Europe PMC as well. Currently there is no automated check to identify deleted preprint records.

Enriching preprints with relevant resources

Along with aggregating preprint content, Europe PMC enriches them in a number of ways. Preprints in Europe PMC are linked to authors, citations, funding, data, reviews, impact metrics, and other resources.

ORCID iDs

ORCID iDs are unique researcher identifiers that help disambiguate academic authors. Information about preprint author ORCID iDs is retrieved programmatically from ORCID. Europe PMC also provides an ORCID linking tool (https://europepmc.org/orcid/import). This tool allows authors to connect articles and preprints in Europe PMC to their ORCID records.

Reviews and evaluations

Preprints in Europe PMC are linked to preprint evaluations from various preprint review communities (Table 4). Preprint review information is retrieved daily from Crossref, and two preprint peer review aggregation platforms - Sciety (https://sciety.org/) and Early Evidence Base (https://eeb.embo.org/).

Preprint review provider	**Metadata source**	**Metadata format**
Arcadia Science	Sciety	DocMaps file

ASAPbio crowd review	Sciety	DocMaps file
Biophysics Colab	Sciety	DocMaps file
eLife	Early Evidence Base	DocMaps file
EMBO Press	Early Evidence Base	DocMaps file
Microbiology Society	Crossref	DOI-associated metadata
Life Science Editors	Sciety	DocMaps file
Life Science Editors Foundation	Sciety	DocMaps file
2019 Novel Coronavirus Research Compendium (NCRC)	Sciety	DocMaps file
NIHR Open Research	Crossref	DOI-associated metadata
Peer Community In	Early Evidence Base	DocMaps file
PeerJ	Sciety	DocMaps file
PeerRef	Early Evidence Base	DocMaps file
preLights	Sciety	DocMaps file
PREreview	Sciety	DocMaps file
Publons	Crossref	DOI-associated metadata
Qeios	Crossref	DOI-associated metadata
Rapid Reviews Infectious Diseases	Sciety	DocMaps file
Review Commons	Early Evidence Base	DocMaps file

ScienceOpen	Crossref	DOI-associated metadata

Table 4. Preprint review groups with review information in Europe PMC.

The Crossref REST API is queried daily to identify review materials by making use of the 'from-index-date' and 'until-index-date' filters to check for new and updated works in Crossref, combined with filtering for works of type 'peer-review'. DOIs of reviewed preprints are matched to preprint records in Europe PMC.

Preprint review metadata is provided in DocMap format by Sciety and Early Evidence Base and is fetched using the following API calls: https://sciety.org/docmaps/v1/index?updatedAfter=[date] and https://eeb.embo.org/api/v2/docmap/[startdate]/[enddate]/[page] from Sciety and Early Evidence Base, respectively. New and updated DocMap files are converted to XML documents using an open source parser [46] developed by Europe PMC, and DOIs of reviewed preprints are matched to preprint records in Europe PMC.

Funding information

Funding information for preprints is retrieved via the Crossref API where available. Mentions of the Europe PMC funder grants are text-mined from the Acknowledgements and Funding sections of full text preprints.

Biological annotations

Preprints in Europe PMC are enriched with annotated biological concepts. Annotation types include organisms, data accessions, or experimental methods. Those concepts are identified in abstracts and full text, where available, by the Europe PMC text-mining pipeline [47].

External links

Preprints in Europe PMC are linked to relevant free external resources, for example biological data records or alternative metrics, through the External Links service (https://europepmc.org/LabsLink).

References

Where possible, articles within reference lists of the full text preprints are matched and hyperlinked to corresponding records in Europe PMC.

Data collection and analysis

Data collection and analysis were performed using the R programming language, making use of the europepmc R package [48], which retrieves data from the publicly accessible Europe PMC APIs. A full description of the methodology documented in an R Markdown file and underlying datasets are available in Zenodo [49].

Results

Including preprints in Europe PMC search

From the start, it was important for Europe PMC to offer a single search solution for life science preprints across many different preprint platforms. As of 4 April 2024, Europe PMC provides access to abstracts and metadata for 761,123 life science preprints from 32 preprint platforms (Table 1).

Since the launch of the Europe PMC project to index preprints in 2018 [34], the number of preprint servers indexed in Europe PMC grew threefold. We continue to monitor the preprint landscape for new platforms that can be added as trusted preprint content sources. It is crucial that preprint content added to Europe PMC conforms to accepted scientific

standards. Therefore, we have adopted a set of indexing guidelines for preprint servers (https://europepmc.org/Preprints#preprint-criteria). To be included in Europe PMC, servers must have a public statement on screening, plagiarism, misconduct, and competing interests. Preprint servers should have no barriers to access the full text for all preprints. This covers subscriptions or sign-in screens. To ensure transparency and improve discoverability, Europe PMC requires a minimal set of metadata. This includes title, abstract, publication date, and author names. Metadata also has to contain a preprint identifier, such as a DOI (digital object identifier). In addition, there are some technical considerations for indexing preprint content. Europe PMC requires access to selected metadata in a machine-readable format. The server has to have a certain volume of preprints posted, with at least 30 preprints available. As a life science database, Europe PMC only includes servers with a significant proportion of life science-related content.

Preprints are included in Europe PMC search results alongside journal publications. If several preprint versions exist in Europe PMC, only the latest version is searchable and shown in search results. Users can access earlier versions through version history on the preprint page (see Table 3 for information on preprint version availability in Europe PMC). When available, the full text of preprints is also included in search. This broadens search results and surfaces preprints that contain search terms not found in the abstracts or metadata. As of 4 April 2024, the full text for 88,831 preprints is searchable in Europe PMC.

Europe PMC also offers advanced search options for preprints [35, 50]. This includes Boolean and syntax searches. For example, it is possible to filter the search to return preprints only (Fig 3A). Additional filters help find preprints that have been published as a journal article (Fig 3B), or have an associated review or evaluation (Fig 3C). It is possible to save regular searches and receive email or RSS alerts for new search results. This allows users to keep up with new preprints in the field.

341 **Fig 3. Europe PMC search filters for preprints, journal-published preprints, and**
342 **reviewed preprints.** Screenshot of the Europe PMC search results shown for a (*Single-cell*
343 *RNA expression profiling of ACE2*) query with the Preprints filter on (A). When the 'Preprints'
344 filter is selected, additional sub-filters are displayed: 'Journal published' filter (B) for preprints
345 with an associated journal published version and 'Reviewed' filter (C) for preprints with
346 associated reviews and evaluations.

347 Tracking changes to preprints

348 Preprints in Europe PMC are clearly differentiated from peer-reviewed articles. Preprints are
349 labelled with a green 'Preprint' tag, both in search results (Fig 4A), and on the preprint page
350 (Fig 4B). The preprint page also displays a yellow banner indicating to the user that the
351 record they are viewing is a preprint (Fig 4C). If available, the preprint review status is also
352 shown in the same banner. Hovering over the question mark sign in the banner brings up
353 additional clarification about preprints (Fig 4D).

354 **Fig 4. Display of preprints in search results (top) and on the preprint record page**
355 **(bottom) in Europe PMC.** The top figure displays a screenshot of the Europe PMC search
356 results page for a (*cognitive decline in TDP-43 proteinopathy*) query. The bottom figure
357 displays a screenshot of the Europe PMC preprint page for the PPR596760 record. (A)
358 Preprints in search results are labelled with a green 'Preprint' tag that also displays version
359 number. (B) The green 'Preprint' tag and (C) a yellow banner indicate that the record being
360 viewed is a preprint. (D) Hovering over the question mark in the banner evokes a pop-up
361 with additional information about preprints. (E) Preprint version history with hyperlinks and
362 publication dates for previous versions.

363 An important feature of preprints is that they can be revised by posting a new version. This
364 allows researchers to correct, update, and improve their article. Europe PMC records version
365 information. Version number is indicated in the green 'Preprint' tag on the search results and

366 preprint page. All preprint versions can be accessed from the ‘Preprint version history’ drop
367 down (Fig 4E). However, version information is only available for some preprints indexed in
368 Europe PMC (Table 3). Preprint servers have different approaches to recording new preprint
369 versions. Some create a unique identifier for each preprint version. Others, notably bioRxiv,
370 medRxiv and SSRN, register a single identifier. In the latter case, the metadata is
371 continuously updated to reflect the most recent version of the preprint. Europe PMC detects
372 new preprint versions through an automated process. This process relies on preprint
373 versions having their own unique persistent identifiers. For preprints with a single identifier,
374 metadata in Europe PMC is overwritten and reflects the latest version. In such cases no
375 version history can be recorded in Europe PMC.

376 As of 8 April 2024 version information can be retrieved for 423,580 preprints in Europe PMC.
377 Of those, 42,458 preprints are known to have more than one version. The largest number
378 recorded in Europe PMC for a single record is 64 versions. Version distribution for preprints
379 in Europe PMC is shown in Figure 5.

380 **Fig 5. Number of preprint records that have more than one version in Europe PMC.**
381 This graph displays the number of preprints for which the latest version available is version 2
382 or higher. The data presented in this graph excludes preprints from servers that issue a
383 single identifier for all versions. Accessed 8 April 2024.

384 Preprints are reciprocally linked to the corresponding journal publications in Europe PMC.
385 The link to the journal article is displayed in the yellow warning banner for all preprint
386 versions (Fig 6A). In turn, the blue information banner on the journal article page provides a
387 link to the latest preprint version (Fig 6B). It is possible to find preprint–journal article pairs in
388 Europe PMC using HAS_PUBLISHED_VERSION:y and HAS_PREPRINT:y searches. As of
389 4 April 2024, 269,479 of preprints have an associated journal published version in Europe
390 PMC. The share of preprints that have been later published in a journal increases with time.
391 For example, 22% preprints posted in 2023 have been published, in comparison to 62% of

392 preprints posted in 2013 (Fig 6C). The reason for this increase is likely caused by delays
393 associated with the journal publishing process. The actual proportion of published preprints
394 in Europe PMC is likely higher than indicated because links from preprints to journal articles
395 can only be established for publications that have a PubMed ID (PMID) and are indexed in
396 Europe PMC.

397 **Fig 6. Preprint–journal article links in Europe PMC.** Screenshots of the information
398 banner on the preprint page with a link to the published journal article (A) and corresponding
399 banner on the journal article page with a link to the latest preprint version (B). A graph
400 showing the percentage of preprints with an associated journal published version in Europe
401 PMC between 2013-2023 (C). The data presented here includes Open Research Platforms,
402 such as F1000 Research, Access Microbiology, etc. Accessed 8 April 2024.

403 Posting a new version of the preprint can allow authors to correct or improve the manuscript.
404 In some cases a preprint may need to be fully suppressed from the preprint server. This can
405 happen due to incorrect data or its interpretation, authorship disputes, legal issues, or as a
406 result of erroneous posting. Preprint platforms have different policies regarding permanence
407 of the preprint content (https://asapbio.org/preprint-servers). There are two main options
408 available to authors: preprint withdrawals and preprint removals. In the case of a withdrawal,
409 the preprint itself is still accessible. It is supplemented with a new version, a withdrawal
410 notice. The withdrawal notice explains that the preprint should not be considered part of the
411 scientific record. This is equivalent to a retraction for journal published articles. In the case of
412 a removal, all preprint versions are deleted and the content is no longer accessible. In some
413 cases, a removal notice replaces the preprint itself [44].

414 In Europe PMC, removed and withdrawn preprints are clearly labelled with a red warning
415 banner (Fig 7A and 7B). Withdrawal banner displays a link to the preprint server for more
416 information. The same banner is also present on earlier versions if those exist in Europe

PMC (Fig 7C). Withdrawn and removed preprints in Europe PMC can be found using PUB_TYPE:"preprint-withdrawal" and PUB_TYPE:"preprint-removal" searches, respectively.

Fig 7. Display or removed and withdrawn preprints in Europe PMC. Warning banners for removed preprint (A), withdrawn preprint (B), and preprint version preceding a withdrawn preprint (C).

Europe PMC determines whether a preprint has been withdrawn or removed using a semi-automated process. This process is a part of the pipeline indexing preprint full text. Withdrawal and removal notices are typically only a few sentences long. Therefore they can be identified based on the length of the preprint full text. Such short preprint records are flagged by the Europe PMC plus system. They are then manually checked and tagged with the appropriate article type. As of 4 April 2024, 56 withdrawn or removed full text preprint records are available in Europe PMC. This comprises 0.1% of all full text preprints. The number does not include preprints that have been deleted from the server entirely, where the preprint URL leads to a 404 page. When such instances are reported to the Europe PMC team, those records are fully removed from Europe PMC to reflect the server's approach.

To help users check for updates to preprint records, Europe PMC has developed the Article Status Monitor (http://europepmc.org/ArticleStatusMonitor). This tool identifies most recent preprint versions, journal published versions, withdrawals or removals. Users can submit a preprint DOI or PPRID to view and export updates to the record and access the latest version.

Preprint status changes can be tracked programmatically, with the Status Update Search module of the Articles RESTful API (http://europepmc.org/RestfulWebService#!/Europe32PMC32Articles32RESTful32API/statusUpdateSearch).

Text-mining and programmatic access for preprints

To support programmatic access to preprints Europe PMC offers freely available APIs. This includes RESTful and SOAP web services with output in XML, JSON or Dublin Core formats. Preprint metadata from preprint servers is enriched with data from other sources. The core API response includes information about funding, reviews, author ORCID iDs, or journal published versions for preprints in Europe PMC. The search syntax reference (https://europepmc.org/searchsyntax) can be used to create advanced programmatic preprint queries, for example to find refereed preprints supported by a particular funding agency.

Preprint abstracts as well as full text, open access preprints, can be downloaded in bulk on the Europe PMC FTP site. Full text and metadata is shared in JATS XML format. This provides a unique set of documents for text-mining and machine-learning purposes.

Europe PMC uses text mining to surface evidence presented in preprints. This process annotates preprints with relevant biological concepts. Example annotations include experimental methods, organisms, diseases, or data accessions. Europe PMC identifies these concepts using text-mining algorithms [51]. The text-mining process covers preprint abstracts and open access full text. Annotations can be accessed using the freely available Europe PMC Annotations API (https://europepmc.org/AnnotationsApi). Annotations can also be highlighted in the text using the SciLite tool [47]. This is possible for all preprint abstracts, and full text preprints with a license that allows derivatives of the work. This is restricted to full text preprints with a CC-BY, CC-BY-NC or CC0 license. Individual annotations are linked to records in public databases or controlled ontologies. For example, a gene ontology term found in the text will be linked to a corresponding record in the Gene Ontology database [52]. This helps foster connections between scientific literature and data.

465 # Incorporating preprints into Europe PMC infrastructure

466 Europe PMC aims to establish preprints as first class research outputs. To this end preprints
467 are linked to relevant scholarly resources. All of the relevant content is summarised on the
468 preprint page. This includes the preprint abstract and full text where available (Fig 8A).
469 Europe PMC also provides tools to incorporate preprints into research workflows. Users can
470 cite a preprint using the 'Get citation' feature. Preprint authors can add the preprint to their
471 ORCID profile using the 'Claim to ORCID' option. Those tools are available within the
472 navigation panel on the right (Fig 8B). The navigation panel on the left (Fig 8C) provides
473 access to related content. This covers data behind the paper, peer review materials, or
474 impact metrics. These resources are available within corresponding sections on the preprint
475 page.

476 **Fig 8**. **Preprint record display in Europe PMC.** Screenshot of the preprint page for the
477 PPR130125 record. (A) Preprint abstract, as well as figures and full text, where available,
478 are accessible from the navigation panel on the left. The same panel provides access to
479 additional sections with links to related content (B), such as 'Citations & impact', 'Data',
480 'Reviews and evaluations', and 'Funding'. The navigation panel on the right provides options
481 to 'Get citation', 'Open PDF', or 'Claim to ORCID' (C).

482 The 'Citations and impact' section displays traditional citations and alternative metrics. The
483 number of supporting or contrasting citations available via the 'scite' tool [53] is shown where
484 possible. Users can also find data citations to the preprint, and the number of preprint
485 recommendations by various groups in this section.

486 The 'Data' section summarises all research data associated with the preprint. It lists data
487 citations, either as data accessions or data DOIs for over 60 different life science databases
488 [54]. Europe PMC identifies citations of data in the text of the preprint using text-mining.
489 Those citations are then linked to the corresponding data record in the external database.

The ‘Data’ section also displays citations from database records to the preprint. For example, if a preprint has been curated by a knowledgebase, or if data behind the preprint was deposited to a dedicated archive.

The ‘Funding’ section contains funding information for preprints in Europe PMC. Where available, grant details include funder name and grant identifier. Additional information is displayed for grants awarded by Europe PMC funders (https://europepmc.org/Funders/). This covers grant award title, principal investigator name and affiliation. Data provided to Europe PMC directly by the Europe PMC funders is freely accessible via the Grant Finder tool (https://europepmc.org/grantfinder).

Linking preprints to reviews and evaluations

The process of posting preprints is decoupled from journal-organised peer review. This allows various preprint review initiatives to experiment with new forms of feedback. As the number of such initiatives grows, it becomes increasingly important to connect preprints with existing reviews that are often scattered across different platforms [55].

For some time Europe PMC has linked peer reviews to preprints via the external links scheme (https://europepmc.org/LabsLink). This approach enabled readers to easily find preprint evaluations but provided limited context, lacking information about review type, date, or name of the reviewer.

To enrich preprint review information in Europe PMC we collaborated with two preprint review aggregation platforms, Sciety and Early Evidence Base. To represent different review systems these platforms have adopted the DocMaps framework [56]. DocMaps is a community-endorsed open-source project led by Knowledge Futures. It uses a standard vocabulary to record steps in the review process. By incorporating DocMaps into Europe PMC a more comprehensive display of preprint review activity could be shown. Some preprint evaluation groups are not yet included in Sciety or Early Evidence Base. Therefore

515 we created a pipeline to pull preprint review metadata available in Crossref. Altogether
516 Europe PMC collates review information from 20 platforms and communities (Table 4). As of
517 4 April 2024 there are 12,209 reviewed preprints in Europe PMC, which can be found using
518 the (HAS_VERSION_EVALUATIONS:Y) search syntax.

519 Preprint review status is disclosed in the yellow warning banner (Figure 9A and 9B). For
520 earlier preprint versions it is also shown in the version history (Figure 9C). The banner
521 provides a link to the 'Reviews and evaluations' section on the preprint page. This section
522 collates review information available in Europe PMC. Readers can explore the review
523 timeline with review activity grouped by platform. For each review Europe PMC provides a
524 link to the full assessment. Where available, Europe PMC shows evaluation date, review title
525 or type, and evaluator name (Figure 9D).

526 **Fig 9. Linking reviews to preprints in Europe PMC.** Information banner for a reviewed
527 preprint (A) and a preprint posted on an Open Research platform (B). (C) Information banner
528 and version history indicate that reviews and evaluations are available for a later version of a
529 preprint. (D) Screenshot of the 'Reviews and evaluations' section for the PPR279548 record.

530 Discussion

531 Benefits of including preprints in Europe PMC

532 Europe PMC started indexing preprints in 2018. At that time our main goal was to make
533 preprints more discoverable and enable their inclusion into research workflows [34]. Since
534 then, further benefits of indexing preprints in Europe PMC became apparent. This includes
535 opportunities for meta research as well as new publishing and review models.

Accelerating discovery of preprints

Europe PMC offers a single search for life science journal articles and preprints. This improves preprint visibility, aided by preprints' default inclusion in search results. Indexing the full text of preprints also increases discoverability. Europe PMC offers full text search for the COVID-19 and Europe PMC funder subsets. This helps surface preprints with search terms found beyond the preprint title or abstract.

Continued proliferation of preprint servers can make it difficult to collate all relevant research for literature reviews, as users may not be aware of emerging preprint platforms. Europe PMC enables comprehensive reviews of preprinted literature. Currently Europe PMC aggregates preprints from over 30 different servers. We continue to increase preprint coverage through a regular monitoring process and welcome inclusion requests directly from preprint servers. To support users who want to follow new preprints in their field, Europe PMC provides saved search and email alert features [50].

Supporting innovative publishing and peer review models

Preprints provide a way to experiment with new forms of communication and innovative review models. Currently, there is a great variety of publishing workflows that incorporate preprints. Journals may offer concurrent submission, use journal-agnostic review services, or opt for post-publication review [57-58]. Diversity in this space provides an opportunity to redefine scholarly communication, but it can be difficult to adapt existing systems to new approaches.

Europe PMC integrates preprints with associated versions, reviews, and published journal articles. This brings together content across different platforms and provides a complete picture of the publishing process. Surfacing preprints and preprint reviews helps recognise and credit preprint authors and referees. Europe PMC works closely with the preprint community to better understand their needs [44]. This allows us to build more flexible tools

561 that accommodate evolving practices. For example, Europe PMC displays modified banners
562 for preprints from Open Research Platforms (Figure 9B). Those platforms coordinate peer
563 review immediately on submission and therefore do not fit the standard definition of a
564 preprint server.

565 Building trust in preprints

566 Since preprints lack the traditional endorsement provided by journal peer review, there is a
567 growing need for the scientific community to find new ways to evaluate preprinted research.
568 Indexing preprints in Europe PMC helps build trust and support continued preprint adoption
569 in many different ways.

570 Collating information about available reviews and recommendations on the preprint page
571 helps readers to assess findings reported in preprints. Users that want to limit their search to
572 preprints that have been evaluated by experts can do so using a dedicated filter.

573 Adopting indexing criteria for preprint servers, such as a requirement for a public statement
574 on screening, plagiarism, and misconduct policies, ensures that preprints in Europe PMC
575 conform to accepted scientific standards.

576 Comprehensive, high quality metadata can improve transparency and provide credibility for
577 preprints. Preprints in Europe PMC are clearly labelled and distinguishable from journal
578 articles. Version information is displayed where available and can be easily accessed from
579 the ‘Preprint version history’ menu. Preprints are also reciprocally linked to the
580 corresponding peer reviewed publications.

581 As concerns regarding preprints’ permanence can erode public trust in science, Europe
582 PMC follows best practices for handling preprint withdrawals and removals. In addition to
583 providing clear notices for affected records, Europe PMC offers a way to filter for withdrawn
584 and removed preprints by using dedicated search syntax. Users can also get notifications in
585 the case of withdrawal or removal of a particular record through email alerts or Article Status

Monitor tool. To support other tools and services that track preprint updates, for example citation managers, information about preprint withdrawals and removals can also be retrieved programmatically using the Article Status API (https://europepmc.org/RestfulWebService#!/Europe32PMC32Articles32RESTful32API/statusUpdateSearch).

Finally, access to underlying data, information on authors, affiliations, and funding can serve as indicators of trust for preprints. Linking preprints in Europe PMC to other research outputs provides a way to assess the evidence presented in preprints in the absence of journal peer review.

Integrating preprints into research workflows

The inclusion of preprints in Europe PMC is an important step in establishing preprints as first class research outputs and integrating preprints into the scholarly infrastructure. Europe PMC tools enable inclusion of preprints into research workflows, such as article citation, funding reporting, and credit and attribution.

When working with preprints, researchers need to know which version to cite, what changes were made to the published version, and whether conclusions still stand after a withdrawal [44]. The Europe PMC Article Status Monitor tool helps users to check for various updates to preprint records, while Europe PMC ‘Get citation’ feature makes it easy to cite preprints or add them to a reference manager tool.

At present, several research funding organisations take preprints into consideration in grant award decisions [19], and many preprints provide information on funding and support. Europe PMC collates funding information for preprints in a dedicated ‘Funding’ section on the preprint page. Europe PMC also enables funders to monitor open data compliance associated with their funding by linking preprints to underlying data, such as data DOIs and accessions text-mined from preprint records.

Supporting preprints inclusion in meta-analyses and systematic reviews

With continued adoption of preprints in the life sciences they are increasingly included in meta-analyses of research literature [59]. Systematic evidence searches require comprehensiveness, transparency and reproducibility [60]. Europe PMC enables preprint inclusion in evidence synthesis through comprehensive coverage of biomedical preprints across multiple platforms, extensive support documentation, and advanced search functionalities.

As a long-standing public service provider that has adopted the Principles of Open Scholarly Infrastructure (POSI) [61], Europe PMC is committed to long term preservation of scholarly records, sustainability, and open access to research literature. This is important for systematic reviewers that rely on lasting availability of content, and transparency in indexing policies and algorithms for reproducible search strategies.

Europe PMC users can create complex, customised queries for preprints using Boolean operators, syntax for field searches, as well as advanced search options. Providing a full text search for preprints in Europe PMC increases recall and allows reviewers to identify specific study types, or limit their search to specific preprint sections, such as Figures, Results, or Methods. Systematic reviewers can export all or a selection of search results in various data formats including XML and CSV.

As a result, Europe PMC adoption as a search engine for preprints in systematic reviews of the research literature has been referenced by numerous studies [62-71].

Providing programmatic access for preprints

While expert meta-analyses are crucial to keep up with the newly published evidence, they are often reliant on several manual, resource-intensive steps [72]. Automated information extraction from a large bibliographic database requires access to programmable interfaces.

For users carrying out bioinformatic studies or literature reviews, Europe PMC offers a single API and a bulk download option for free and open access to preprint abstracts and appropriately licensed full text preprints from many different servers. This makes analysis easier and accelerates research of the preprint literature itself. It also enables creation of new tools and cross-platform integrations. For example, the Europe PMC API powers FlyBase preprint search for preprints relevant to the *Drosophila* community [73] and StemRxiv preprint search from the StemJournal for preprints on stem cell research (https://stemjnl.org/preprint-search).

Crucially, Europe PMC offers one of the largest corpora of preprinted literature in a standard, machine-readable format, providing a unique set of documents for text-mining and machine-learning purposes. In addition, preprints in Europe PMC are incorporated into the knowledge graph and enriched with additional data, including links to journal publications, funding information, etc. High quality open metadata associated with preprints in Europe PMC supports analysis of preprint practices, for example around the impact of peer review, sharing behaviours, the availability of supporting data for preprints, ORCID uptake and so on, at a scale and breadth not previously possible. At present, the full text preprint subset available in Europe PMC has been used for various analyses of research reported in preprints [17, 74-78].

Technical challenges of indexing preprints

Fully realising the potential of preprints by the scholarly community requires a great degree of coordination between servers [44, 79]. Although there are several best practice recommendations jointly developed by the preprint community [44, 80], no standards are currently enforced for preprints. As a preprint aggregator, Europe PMC has faced several technical challenges related to preprint management and variability in metadata standards that exist between preprint platforms.

Distributed access points

A distributed network of preprint servers ensures a degree of autonomy for each preprint community. However, it also creates a challenge of retrieving preprint metadata from different platforms for literature databases, such as Europe PMC. Currently Europe PMC relies on the Crossref API to retrieve preprint metadata and updates to preprint records. Although the majority of the preprint servers included in Europe PMC register DOIs and deposit associated metadata for preprints in Crossref, some do not. This does not constitute an outright barrier to inclusion in Europe PMC, but requires free API access and documentation provided by the server, as well as additional resources from the indexer side to set up separate ingest processes for new content sources.

Limited metadata

Europe PMC requires a minimal set of metadata as part of its preprint indexing guidelines to ensure discoverability of preprints in search results. This includes a persistent identifier, title, abstract, posted date, and author names. However, there is a lot of variability in the scope of metadata deposited by preprint servers to Crossref, and some preprint servers do not supply all of the metadata fields that Europe PMC requires. This prevents their inclusion in Europe PMC.

Divergence of practices and standards

As the preprint ecosystem evolves, the importance of community standards and best practices for preprints becomes more apparent. Aggregating preprints from different sources poses a challenge of accurately representing different approaches. For example, the latest preprint version is equivalent to the peer reviewed article for Open Research platforms and journals adopting the publish-review-curate model [17, 21-24]. Distinguishing these records requires adoption of a basic taxonomy for document versions. Harmonising on these issues will help build a more transparent system.

Lack of machine-readable status updates

Unlike journal articles, preprints are not static in their nature and can often be revised, reviewed, published, withdrawn, or removed. Tracking these changes can be difficult due to a number of reasons.

Handling of preprint versions differs from server to server. Some best practice recommendations call for a unique identifier for each preprint version [44], while Crossref recommends updating the preprint metadata using relations tags for new versions (https://www.crossref.org/documentation/research-nexus/posted-content-includes-preprints/). Where these guidelines are not followed, version history can not be recorded in Europe PMC. Gaps in Crossref metadata deposition lead to missing links between preprints and published articles [81]. While basic automated mechanisms can establish those connections without a metadata flag, they may fail if authors or the title change from preprint to the journal version. Finally, preprint withdrawal or removal statuses are not currently available from a centralised source in a standard machine-readable format. This prevents Europe PMC and other providers from automating updates to the preprint record and requires laborious manual checks.

License limitations

While preprints offer free access for readers, they are not always free to reuse. License limitations can make it difficult to include preprints in large scale analyses. Text-mining requires access to full text in a structured machine-readable format. Preprints are often submitted by the authors as PDF files. Thus there is a need to convert the full text of preprints to a suitable format, such as XML. Regrettably, conversion costs often make it prohibitively expensive.

Europe PMC is able to create a full text preprints corpus in XML format thanks to support from its funders. Our first full text preprint project focused on COVID-19 preprints [36]. This

710 initiative was supported by a joint award from Wellcome, UK Medical Research Council
711 (MRC), Swiss National Science Foundation (SNSF) and endorsed by the Chief Scientist of
712 the World Health Organisation (WHO). The full text for all relevant COVID-19 preprints was
713 converted to XML irrespective of the preprint licence. For preprints with a Creative Commons
714 (CC) license or similar the full text could then be displayed on the Europe PMC website for
715 reading. The preprint full text was also included in the open access subset for bulk
716 downloads and distributed via Europe PMC APIs. Where the current preprint license did not
717 allow reuse, the author was asked to apply the Europe PMC open access licence. The lack
718 of author licence approval resulted in an excess of 27,000 COVID-19 preprints with full text
719 searchable in Europe PMC but not available to read and text mine. This hinders reuse of the
720 full text preprint corpus and reduces its benefits. As a result, when the Europe PMC funder
721 preprint initiative was launched only preprints with Creative Commons licences were
722 selected for inclusion. Due to associated conversion costs this initiative is currently limited to
723 preprints that acknowledge funding from Europe PMC funders.

724 Some servers create XML files for preprints shared on their platform, this includes bioRxiv
725 and medRxiv. However, their license prevents redistribution, precluding Europe PMC from
726 adding those preprints to the collection of open access literature.

727 Future of preprints in Europe PMC

728 Increasing preprint coverage

729 When preprints were first included in Europe PMC in 2018, this was limited to servers that
730 registered Crossref DOIs. Using Crossref provided Europe PMC with a robust centralised
731 process to pull relevant preprint metadata in a standard format. While the majority of the life
732 science preprint servers are current Crossref members, some other servers use alternative
733 DOI registration agencies, including DataCite (e.g. arXiv) or Japan Link Center (e.g. Jxiv). In

the future we hope to extend the Europe PMC preprint pipeline to be able to add preprint metadata from sources other than Crossref.

Enriching preprint metadata

Preprint servers often operate under resource constraints, which can make it difficult to capture comprehensive metadata for every record. One example of limited Crossref metadata for preprints is lack of author affiliations, which are considered desired but not essential metadata by the current guidelines for preprint servers [44].

As Europe PMC already generates XML files for full text preprints we can extract additional information that is not available in Crossref. This allows us to surface affiliations for full text preprints in Europe PMC. As a potential next step we plan to add Research Organization Registry identifiers (ROR IDs) to affiliations for Europe PMC-generated full text preprints. ROR provides open persistent identifiers for research organisations that help disambiguate institutional names and connect them to relevant research outputs.

Including preprints in the citation network

Citation information in Europe PMC is currently retrieved from PubMed and is limited to journal articles. In the future we plan to extend the Europe PMC open citation network to include preprint citations. Potential citation sources include openly available Crossref metadata supplied by publishers, as well as citations determined from the reference lists of full text publications and preprints in Europe PMC.

Expanding links between preprints and reviews

While DocMaps provide a comprehensive schema for capturing preprint review information, currently they are only created for a limited number of review groups. Some platforms that offer preprint review instead choose to register DataCite DOIs for referee reports, for

example F1000 Open Research Platforms. To ensure comprehensive coverage of preprint evaluations we will investigate options for capturing review information currently missing from Europe PMC using the DataCite API (https://api.datacite.org).

Standards development

Europe PMC has played a leading role in preprint community standards development and helped actualize recommendations for metadata practices through a series of stakeholder events [44, 82].

In the future we will continue to work closely with the preprints community on standards for preprint metadata and full text content. Standards for notifications for preprint withdrawals and removals, as well as other metadata events, such as the linking of reviews and evaluations comments to a preprint, are of particular relevance to our work.

Acknowledgements

We thank arXiv and SSRN staff for providing us with a curated list of COVID-19 preprints during the COVID-19 full text preprint initiative.

References

1. Grossmann A, Brembs B. Current market rates for scholarly publishing services [version 2; peer review: 2 approved]. F1000Research. 2021;10: 20. doi: 10.12688/f1000research.27468.2.
2. Björk BC, Solomon D. The publishing delay in scholarly peer-reviewed journals. J Informet. 2013;7: 914–23. doi: 10.1016/j.joi.2013.09.001.
3. Lee KP, Boyd EA, Holroyd-Leduc JM, Bacchetti P, Bero LA. Predictors of publication: characteristics of submitted manuscripts associated with acceptance at major

779 biomedical journals. Med J Aust. 2006;184(12): 621-626. doi: 10.5694/j.1326-
780 5377.2006.tb00418.x.
781 4. Tregenza T. Gender bias in the refereeing process? Trends Ecol Evol. 2002;17(8):
782 349–50. doi: 10.1016/j.molimm.2014.11.005.
783 5. Shen YA, Webster JM, Shoda Y, Fine I. Persistent Underrepresentation of Women's
784 Science in High Profile Journals. bioRxiv [Preprint]. 2018 bioRxiv 275362 [posted
785 2018 Mar 2; revised 2018 Mar 8; cited 2024 April 4]. Available from:
786 https://www.biorxiv.org/content/10.1101/275362v2. doi: 10.1101/275362.
787 6. Murray D, Siler K, Larivière V, Chan WM, Collings AM, Raymond J, et al. Gender and
788 international diversity improves equity in peer review. bioRxiv [Preprint]. 2019 biorXiv
789 400515 [posted 2018 Aug 29; revised 2019 Apr 11; revised 2019 Aug 4; cited 2024
790 Apr 4]. Available from: https://www.biorxiv.org/content/10.1101/400515v3. doi:
791 10.1101/400515.
792 7. Seeber M, Bacchelli A. Does single blind peer review hinder newcomers?
793 Scientometrics. 2017;113(1): 567-585. doi: 10.1007/s11192-017-2264-7.
794 8. Steen RG, Casadevall A, Fang FC. Why has the number of scientific retractions
795 increased? PLOS One. 2013;8(7): e68397. doi: 10.1371/journal.pone.0068397.
796 9. Fang FC, Steen RG, Casadevall A. Misconduct accounts for the majority of retracted
797 scientific publications. Proceedings of the National Academy of Sciences of the
798 United States of America. 2012;109(42): 17028-17033. doi:
799 10.1073/pnas.1212247109.
800 10. Oransky I, Fremes SE, Kurlansky P, Gaudino M. Retractions in medicine: the tip of
801 the iceberg. Eur Heart J. 2021;42(41): 4205-4206. doi: 10.1093/eurheartj/ehab398.
802 11. Berg JM, Bhalla N, Bourne PE, Chalfie M, Drubin DG, Fraser JS, et al. SCIENTIFIC
803 COMMUNITY. Preprints for the life sciences. Science. 2016;352(6288): 899-901. doi:
804 10.1126/science.aaf9133.

12. Sarabipour S, Debat HJ, Emmott E, Burgess SJ, Schwessinger B, Hensel Z. On the value of preprints: An early career researcher perspective. PLOS Biology. 2019;17(2): e3000151. doi: 10.1371/journal.pbio.3000151.
13. Chiarelli A, Johnson R, Pinfield S, Richens E. Preprints and Scholarly Communication: Adoption, Practices, Drivers and Barriers [version 2; peer review: 3 approved, 1 approved with reservations]. F1000Research. 2019;8: 971. doi: 10.12688/f1000research.19619.2.
14. Ginsparg P. Lessons from arXiv's 30 years of information sharing. Nature Reviews Physics. 2021;3: 602–603. doi: 10.1038/s42254-021-00360-z.
15. Confrey EA. The information exchange groups experiment. Publishing Research Quarterly. 1996;12: 37-39. doi: 10.1007/BF02680369.
16. Cobb M. The prehistory of biology preprints: a forgotten experiment from the 1960s. PeerJ Preprints [Preprint]. 2017 PeerJ Preprints 5:e3174v1 [posted 2017 Aug 22; cited 2024 Apr 4]. Available from: https://peerj.com/preprints/3174v1/. doi: 10.7287/peerj.preprints.3174v1.
17. Kirkham JJ, Penfold NC, Murphy F, Boutron I, Ioannidis JP, Polka J, et al. Systematic examination of preprint platforms for use in the medical and biomedical sciences setting. BMJ Open. 2020;10(12): e041849. doi: 10.1136/bmjopen-2020-041849.
18. Tennant J, Bauin S, James S, Kant J. The evolving preprint landscape: Introductory report for the Knowledge Exchange working group on preprints. MetaArXiv [Preprint]. 2018 MetaArXiv [posted 2018 May 17; cited 2024 Apr 4]. Available from: https://osf.io/preprints/metaarxiv/796tu. doi: 10.31222/osf.io/796tu.
19. Penfold NC, Polka JK. Technical and social issues influencing the adoption of preprints in the life sciences. PLOS Genetics. 2020;16(4): e1008565. doi: 10.1371/journal.pgen.1008565.
20. Klebel T, Reichmann S, Polka J, McDowell G, Penfold N, Hindle S, et al. Peer review and preprint policies are unclear at most major journals. PLOS One. 2020;15(10): e0239518. doi: 10.1371/journal.pone.0239518.

21. Marshall H, Efthimiou G. Beyond Access Microbiology: implementing an open research platform for the Microbiology Society. Access Microbiology. 2022;4(7): acmi000369. doi: 10.1099/acmi.0.000369.
22. Pourret O, Ibarra DE. The rise of preprints in earth sciences [version 2; peer review: 2 approved, 1 approved with reservations]. F1000Research. 2023;12: 561. doi: 10.12688/f1000research.133612.2.
23. Chiarelli A, Johnson R, Pinfield S and Richens E. Preprints and Scholarly Communication: An Exploratory Qualitative Study of Adoption, Practices, Drivers and Barriers [version 2; peer review: 3 approved, 1 approved with reservations]. F1000Research. 2019;8: 971. doi: 10.12688/f1000research.19619.2.
24. Eisen MB, Akhmanova A, Behrens TE, Diedrichsen J, Harper DM, Iordanova MD, et al. Scientific Publishing: Peer review without gatekeeping. eLife. 2022;11: e83889. doi: 10.7554/eLife.83889.
25. Gigascience. GigaByte journal joins the 'Publish, Review, Curate' transformation of the publishing paradigm. 2023 Sep 26 [cited 2024 Apr 4]. In: EurekAlert! [Internet]. Available from: https://www.eurekalert.org/news-releases/1002740.
26. Lin J. Preprints growth rate ten times higher than journal articles. 2018 May 31 [cited 2024 Apr 4]. In: Crossref Blog [Internet]. Available from: https://www.crossref.org/blog/preprints-growth-rate-ten-times-higher-than-journal-articles/.
27. Johansson MA, Reich NG, Meyers LA, Lipsitch M. Preprints: An underutilized mechanism to accelerate outbreak science. PLOS Medicine. 2018;15: e1002549. doi: 10.1371/journal.pmed.1002549.
28. Wellcome. Sharing research data and findings relevant to the novel coronavirus (COVID-19) outbreak. 2020 Jan 31 [cited 4 April 2024]. In: Wellcome press releases [Internet]. Available from: https://wellcome.org/press-release/sharing-research-data-and-findings-relevant-novel-coronavirus-ncov-outbreak.

29. WHOSTP. COVID-19 open-access letter from chief scientific advisors to members of the scholarly publishing community. 2020 Mar 13 [cited 4 April 2024]. In: White House Archives [Internet]. Available from: https://trumpwhitehouse.archives.gov/wp-content/uploads/2020/03/COVID19-Open-Access-Letter-from-CSAs.Equivalents-Final.pdf.
30. Waltman L, Pinfield S, Rzayeva N, Henriques SO, Fang Z, Brumberg J, et al. Scholarly communication in times of crisis: The response of the scholarly communication system to the COVID-19 pandemic. Research on Research Institute. Report; 2021 [cited 2024 Apr 4]. Database: figshare [Internet]. Available from: https://doi.org/10.6084/m9.figshare.17125394.v1.
31. Microsoft. Next Steps for Microsoft Academic – Expanding into New Horizons. 2021 May 4 [cited 2024 Apr 4]. In: Microsoft Academic blog [Internet]. Available from: https://www.microsoft.com/en-us/research/project/academic/articles/microsoft-academic-to-expand-horizons-with-community-driven-approach/.
32. Chan Zuckerberg Initiative Science. Meta Transition. 2021 Oct 28 [cited 2024 Apr 4]. In: Medium [Internet]. Available from: https://cziscience.medium.com/meta-transition-5f66b1fae475.
33. Brietzke E, Gomes FA, Gerchman F, Freire RCR. Should systematic reviews and meta-analysis include data from preprints? Trends in Psychiatry and Psychotherapy. 2023;45: e20210324. doi: 10.47626/2237-6089-2021-0324.
34. Europe PMC Team. Preprints in Europe PMC: reducing friction for discoverability. 2018 Jul 11 [cited 2024 Apr 4]. In: Europe PMC News Blog [Internet]. Available from: https://blog.europepmc.org/2018/07/preprints-in-europe-pmc-reducing-friction-for-discoverability.html
35. Rosonovski S, Levchenko M, Bhatnagar R, Chandrasekaran U, Faulk L, Hassan I, et al. Europe PMC in 2023. Nucleic Acids Research. 2024;52(D1): D1668-D1676. doi: 10.1093/nar/gkad1085.

887 36. Europe PMC Team. Europe PMC to include the full text of COVID-19 preprints. 2020
888 Jun 8 [cited 2024 Apr 4]. In: Europe PMC News Blog [Internet]. Available from:
889 https://blog.europepmc.org/2020/06/europe-pmc-to-include-the-full-text-of-covid-19-
890 preprints.html
891 37. Europe PMC Team. Over 15,300 full text COVID-19 now available in Europe PMC.
892 2021 Feb 9 [cited 2024 Apr 4]. In: Europe PMC News Blog [Internet]. Available from:
893 https://blog.europepmc.org/2021/02/over-15300-full-text-covid-19-now-available-in-
894 europe-pmc.html
895 38. Scheinfeld L. LitCovid, iSearch COVID-19 portfolio, and COVID-19 Global literature
896 on coronavirus disease. Journal of the Medical Library Association. 2022;110(2):
897 279-280. doi: 10.5195/jmla.2022.1274.
898 39. Wang LL, Lo K, Chandrasekhar Y, Reas R, Yang J, Burdick D, et al. CORD-19: The
899 COVID-19 Open Research Dataset. arXiv:2004.10706v4 [Preprint]. 2020 [cited 2024
900 Ap 4]. Available from: https://arxiv.org/abs/2004.10706.
901 40. Europe PMC Team. Europe PMC improves discoverability of preprints. 2022 May 3
902 [cited 2024 Apr 4]. In: Europe PMC News Blog [Internet]. Available from:
903 https://blog.europepmc.org/2022/05/europe-pmc-improves-discoverability-of-
904 preprints.html
905 41. Funk K, Zayas-Cabán T, Beck J. Phase 1 of the National Institutes of Health Preprint
906 Pilot: Testing the viability of making preprints discoverable in PubMed Central and
907 PubMed. bioRxiv [Preprint]. 2022 bioRxiv 520156 [posted 2022 Dec 13; revised 2024
908 Jan 18; cited 2024 Apr 4]. Available from:
909 https://www.biorxiv.org/content/10.1101/2022.12.12.520156v2. doi:
910 10.1101/2022.12.12.520156.
911 42. Europe PMC team. Crossref-preprints-importer (Version 1.0.0); 2023 [cited 2024 Apr
912 4]. In: Gitlab [Internet]. Available from https://gitlab.ebi.ac.uk/literature-
913 services/public-projects/crossref-preprints-importer.

43. Beck J, Ferguson CA, Funk K, Hanson B, Harrison M, Ide-Smith M, et al. Building trust in preprints: recommendations for servers and other stakeholders. OSF Preprints [Preprint]. 2020 OSF Preprints [posted 2020 Jul 21; cited 2024 April 4]. doi:10.31219/osf.io/8dn4w.
44. Europe PMC Team. Transparency for preprints: handling withdrawals and removals. 2021 Dec 1 [cited 2024 Apr 4]. In: Europe PMC News Blog [Internet]. Available from: https://blog.europepmc.org/2021/12/transparency-for-preprints-handling-withdrawals-and-removals.html
45. Hamelers A, Parkin M. A full text collection of COVID-19 preprints in Europe PMC using JATS XML. 2021 [cited 2024 Apr 4]. In: Journal Article Tag Suite Conference (JATS-Con) Proceedings 2020/2021 [Internet]. Bethesda (MD): National Center for Biotechnology Information (US); 2021. Available from: https://www.ncbi.nlm.nih.gov/books/NBK569517/?report=classic
46. Europe PMC team. Docmap-parser (Version 0.1.2-SNAPSHOT); 2023 [cited 2024 Apr 4]. In: Gitlab [Internet]. Available from https://gitlab.ebi.ac.uk/literature-services/public-projects/docmap-parser.
47. Venkatesan A, Kim JH, Talo F, Ide-Smith M, Gobeill J, Carter J, et al. SciLite: a platform for displaying text-mined annotations as a means to link research articles with biological data [version 2; peer review: 2 approved, 1 approved with reservations]. Wellcome Open Res. 2017;1: 25 doi: 10.12688/wellcomeopenres.10210.2.
48. Jahn N. europepmc: R Interface to the Europe PubMed Central RESTful Web Service (R package version 0.4.3); 2024 [cited 2024 Apr 4]. In: Github [Internet]. Available from: https://github.com/ropensci/europepmc/.
49. Parkin M. R Markdown script to accompany the paper 'Enabling preprint discovery, evaluation, and analysis with Europe PMC'; 2024 [cited 2024 Apr 10]. Database: Zenodo [Internet]. Available from: 10.5281/zenodo.10711385.

941 50. Rosonovski S, Levchenko M, Ide-Smith M, Faulk L, Harrison M, McEntyre J.
942 Searching and Evaluating Publications and Preprints Using Europe PMC. Current
943 Protocols. 2023;3(3): e694. doi: 10.1002/cpz1.694.
944 51. McEntyre JR, Ananiadou S, Andrews S, Black WJ, Boulderstone R, Buttery P, et al.
945 UKPMC: a full text article resource for the life sciences. Nucleic Acids Research.
946 2011;39(Database issue): D58-65. doi: 10.1093/nar/gkq1063.
947 52. The Gene Ontology Consortium. The Gene Ontology knowledgebase in 2023.
948 Genetics. 2023;224(1): iyad031. doi: 10.1093/genetics/iyad031
949 53. Nicholson JM, Mordaunt M, Lopez P, Uppala A, Rosati D, Rodrigues NP, et al. scite:
950 A smart citation index that displays the context of citations and classifies their intent
951 using deep learning. Quantitative Science Studies. 2021;2(3): 882–898. doi:
952 10.1162/qss_a_00146.
953 54. Ferguson C, Araújo D, Faulk L, Gou Y, Hamelers A, Huang Z, et al. Europe PMC in
954 2020. Nucleic Acids Research. 2021;49(D1): D1507-D1514. doi:
955 10.1093/nar/gkaa994.
956 55. Avissar-Whiting M, Belliard F, Bertozzi SM, Brand A, Brown K, Clément-Stoneham
957 G, et al. Advancing the culture of peer review with preprints. OSF Preprints [Preprint].
958 2023 OSF Preprints [posted 2023 Apr 3; revised 2024 Feb 14; cited 2024 April 4].
959 Available from: https://osf.io/preprints/osf/cht8p. doi: 10.31219/osf.io/cht8p.
960 56. McDowell GS, Polka JK, Ross-Hellauer T, Stein G. The DocMaps Framework for
961 representing assertions on research products in an extensible, machine-readable,
962 and discoverable format. bioRxiv [Preprint]. 2021 bioRxiv [posted 2021 Jul 15; cited
963 2024 April 4]. Available from:
964 https://www.biorxiv.org/content/10.1101/2021.07.13.452204v1. doi:
965 10.1101/2021.07.13.452204.
966 57. Sever R, Roeder T, Hindle S, Sussman L, Black KJ, Argentine J, et al. bioRxiv: the
967 preprint server for biology. bioRxiv [Preprint]. 2019 bioRxiv 833400 [posted 2019 Nov

968 6; cited 2024 April 4]. Available from:
969 https://www.biorxiv.org/content/10.1101/833400v1. doi: 10.1101/833400.
970 58. Irawan DE, Zahroh H, Puebla I. Preprints as a driver of open science: Opportunities
971 for Southeast Asia. Frontiers in Research Metrics and Analytics. 2022;7: 992942. doi:
972 10.3389/frma.2022.992942.
973 59. Clyne B, Walsh KA, O'Murchu E, Sharp MK, Comber L, O' Brien KK, et al. Using
974 preprints in evidence synthesis: Commentary on experience during the COVID-19
975 pandemic. Journal of Clinical Epidemiology. 2021;138: 203-210. doi:
976 10.1016/j.jclinepi.2021.05.010.
977 60. Gusenbauer M, Haddaway NR. Which academic search systems are suitable for
978 systematic reviews or meta-analyses? Evaluating retrieval qualities of Google
979 Scholar, PubMed, and 26 other resources. Research Synthesis Methods. 2020;11(2):
980 181-217. doi: 10.1002/jrsm.1378.
981 61. Europe PMC Team. Europe PMC adopts the Principles of Open Scholarly
982 Infrastructure. 2022 Feb 22 [cited 2024 April 4]. In: Europe PMC News Blog
983 [Internet]. Available from: https://blog.europepmc.org/2022/02/europe-pmc-adopts-
984 the-principles-of-open-scholarly-infrastructure.html
985 62. Griessbach A, Schönenberger CM, Taji Heravi A, Gloy V, Agarwal A, Hallenberger
986 TJ, et al. Characteristics, Progression, and Output of Randomized Platform Trials: A
987 Systematic Review. JAMA Network Open. 2024;7(3): e243109. doi:
988 10.1001/jamanetworkopen.2024.3109.
989 63. Barreñada L, Ledger A, Dhiman P, Collins G, Wynants L, Verbakel JY, et al. ADNEX
990 risk prediction model for diagnosis of ovarian cancer: systematic review and meta-
991 analysis of external validation studies. BMJ Medicine. 2024;3(1): e000817. doi:
992 10.1136/bmjmed-2023-000817.
993 64. Anand A, Vialard F, Esmail A, Ahmad Khan F, O'Byrne P, Routy JP, et al. Self-tests
994 for COVID-19: What is the evidence? A living systematic review and meta-analysis

995 (2020-2023). PLOS Global Public Health. 2024;4(2): e0002336. doi:
996 10.1371/journal.pgph.0002336.
997 65. Katzenschlager S, Brümmer LE, Schmitz S, Tolle H, Manten K, Gaeddert M, et al.
998 Comparing SARS-CoV-2 antigen-detection rapid diagnostic tests for COVID-19 self-
999 testing/self-sampling with molecular and professional-use tests: a systematic review
1000 and meta-analysis. Scientific Reports. 2023;13(1): 21913. doi: 10.1038/s41598-023-
1001 48892-x.
1002 66. Sulis G, Kim JY, Rodrigue V, Gore G, Peebles A, Ulrich AK, et al. Sex-disaggregated
1003 effectiveness data reporting in COVID-19 vaccine research: a systematic review.
1004 Communications Medicine. 2023;3(1): 69. doi: 10.1038/s43856-023-00297-7.
1005 67. Byambasuren O, Stehlik P, Clark J, Alcorn K, Glasziou P. Effect of covid-19
1006 vaccination on long covid: systematic review. BMJ Medicine. 2023;2(1): e000385.
1007 doi: 10.1136/bmjmed-2022-000385.
1008 68. Bobrovitz N, Ware H, Ma X, Li Z, Hosseini R, Cao C, et al. Protective effectiveness of
1009 previous SARS-CoV-2 infection and hybrid immunity against the omicron variant and
1010 severe disease: a systematic review and meta-regression. The Lancet Infectious
1011 Diseases. 2023;23(5): 556-567. doi: 10.1016/s1473-3099(22)00801-5.
1012 69. Lewis HC, Ware H, Whelan M, Subissi L, Li Z, Ma X, et al. SARS-CoV-2 infection in
1013 Africa: a systematic review and meta-analysis of standardised seroprevalence
1014 studies, from January 2020 to December 2021. BMJ Global Health. 2022;7(8):
1015 e008793. doi: 10.1136/bmjgh-2022-008793.
1016 70. Gammall J, Lai AG. Pan-cancer prognostic genetic mutations and clinicopathological
1017 factors associated with survival outcomes: a systematic review. NPJ Precision
1018 Oncology. 2022;6(1): 27. doi: 10.1038/s41698-022-00269-5.
1019 71. Baker J, Krishnan N, Abroms LC, Berg CJ. The Impact of Tobacco Use on COVID-19
1020 Outcomes: A Systematic Review. Journal of Smoking Cessation. 2022;2022:
1021 5474397. doi: 10.1155/2022/5474397.

72. Bannach-Brown A, Hair K, Bahor Z, Soliman N, Macleod M, Liao J. Technological advances in preclinical meta-research. BMJ Open Science. 2021;5(1): e100131. doi: 10.1136/bmjos-2020-100131.
73. Gramates LS, Agapite J, Attrill H, Calvi BR, Crosby MA, Dos Santos G, et al. FlyBase: a guided tour of highlighted features. Genetics. 2022;220(4): iyac035. doi: 10.1093/genetics/iyac035.
74. Schwab S, Held L. Science after Covid-19: Faster, better, stronger? Significance. 2020;17(4): 8-9. doi: 10.1111/1740-9713.01415.
75. Fraser N, Kramer B. covid19_preprints (Version 58); 2021 [cited 2024 Apr 4]. Database: figshare [Internet]. Available from: https://figshare.com/articles/code/covid19_preprints/12033672.
76. Goh ZH, Coates J, Gruson H. Covidpreprints.com: Automating Website Updates with the europepmc and rAltmetric Packages. 2020 Oct 13 [cited 2024 Apr 4]. In: rOpenSci blog [Internet]. Available from: https://ropensci.org/blog/2020/10/13/covidpreprints/.
77. Ivanova Y, Karapeev G, Butler D, Gadegaard N, Aitchison I, Cannon P, et al. Fluctuations in SDG relevant research output in response to COVID-19; 2020 [cited 2024 Apr 4]. Database: Open Science Framework [Internet]. Available from: https://osf.io/ea37y/.
78. Jahn N. COVID-19 Preprints in Europe PMC: Cumulative daily growth by source; 2020 [cited 2024 Apr 4]. In: Github [Internet]. Available from: https://gist.github.com/njahn82/b598376b2affe47681b61e273e041b55.
79. Malički M, Alperin JP. Four recommendations for improving preprint metadata. 2020 April 8 [cited 2024 Apr 4]. In: Scholcommlab Blog [Internet]. Available from: https://www.scholcommlab.ca/2020/04/08/preprint-recommendations/.
80. Rittman M, Marchant E, Havemann J, Rieger OY, Wrigley AT, Pepe A, et al. Crossref metadata for preprints: Discussions and recommendations. MetaArXiv [Preprint]. 2022 MetaArXiv [posted 2022 Nov 24; revised 2022 Nov 24; revised 2022 Nov 28;

cited 2024 Apr 4]. Available from: https://doi.org/10.31222/osf.io/qzusj. Doi: 10.31222/osf.io/qzusj.

81. Lin J, Ram K. Leaving the house—where preprints go. 2018 Aug 21 [cited 2024 Apr 4]. In: Crossref Blog [Internet]. Available from: https://www.crossref.org/blog/leaving-the-house-where-preprints-go/.

82. Rosonovski S. Building a technological foundation for preprint review. 2023 Oct 25 [cited 2024 Apr 4]. In: Europe PMC News Blog [Internet]. Available from: https://blog.europepmc.org/2023/10/building-a-technological-foundation-for-preprint-review.html.

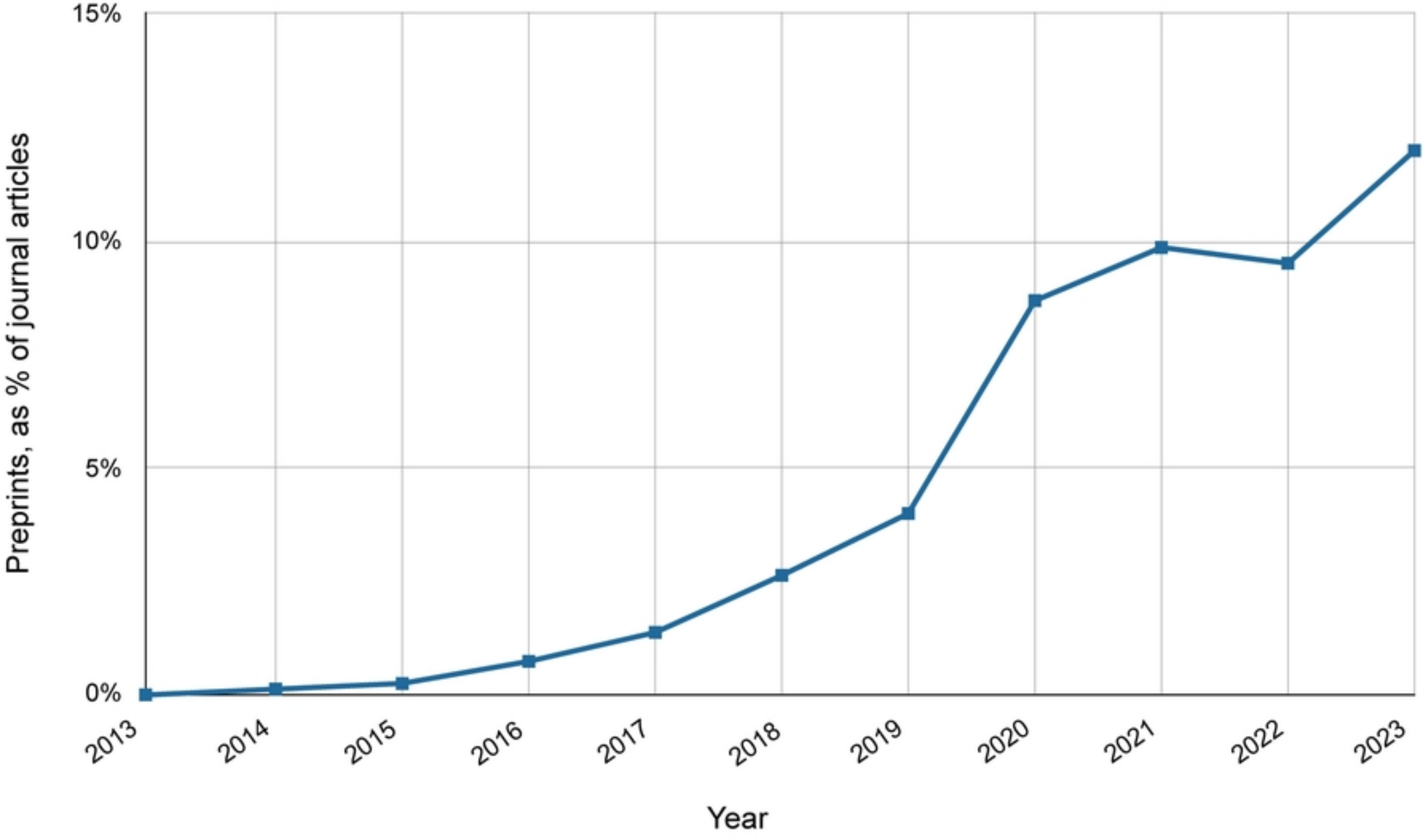

15%
10%
5%
0%
Preprints, as % of journal articles
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022
2023
Year

Fig 1

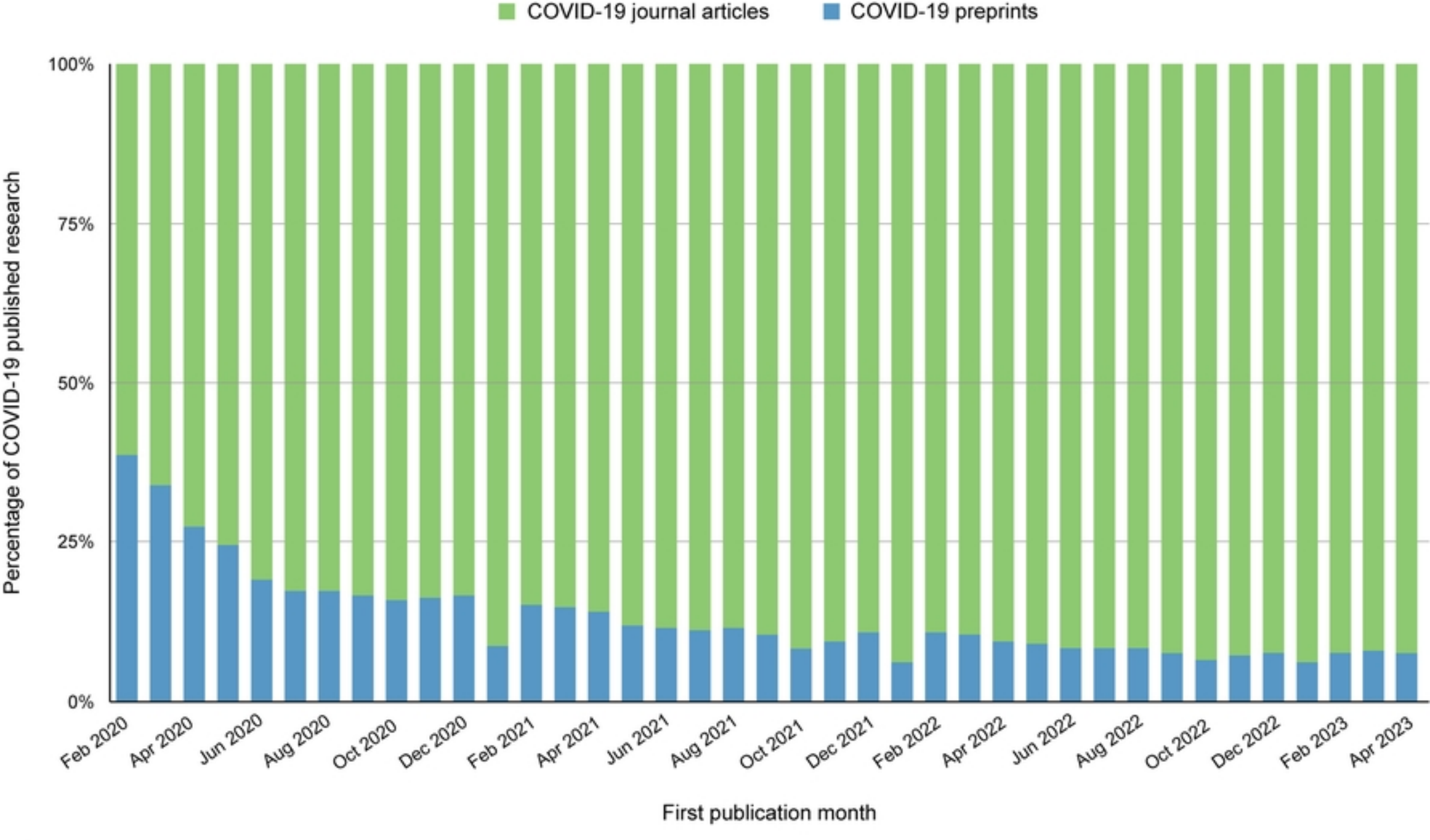

COVID-19 journal articles
COVID-19 preprints
100%
75%
50%
25%
0%
Percentage of COVID-19 published research
Feb 2020
Apr 2020
Jun 2020
Aug 2020
Oct 2020
Dec 2020
Feb 2021
Apr 2021
Jun 2021
Aug 2021
Oct 2021
Dec 2021
Feb 2022
Apr 2022
Jun 2022
Aug 2022
Oct 2022
Dec 2022
Feb 2023
Apr 2023
First publication month

Fig 2

Free full text access

- [ ] Full text in Europe PMC (456)
- [ ] Link to free full text (183)

Type

- [ ] Research articles (1,792)
- [ ] Review articles (885)
- A [x] Preprints (639)
- B [ ] Journal published (320)
- C [ ] Reviewed (21)

1-25 of **639** results

Sort by: Relevance

1 2 3 4 5 Next ...

Single-cell RNA expression profiling of ACE2, the receptor of SARS-CoV-2
Zhao Y, Zhao Z, Wang Y, Zhou Y, Ma Y, Zuo W
bioRxiv, 26 Jan 2020
Cited by: 35 articles | PPR: PPR110216
+ Add to export list
Preprint
Free full text in Europe PMC

Single-cell RNA expression profiling of ACE2, the putative receptor of Wuhan 2019-nCoV, in the nasal tissue
Wu C, Zheng S, Chen Y, Zheng M

Fig 3

1-25 of **661** results

Sort by: Relevance

1 2 3 4 5 Next ...

RNA-binding deficient **TDP-43** drives **cognitive decline** in a mouse model of **TDP-43 proteinopathy**

Necarsulmer J, Simon J, Evangelista B, Chen Y, Tian X, Nafees S, Gonzalez AM, Wang P, Ajit D, Nikolova V, Harper K, Ezzell J, Beltran A, Moy S, Cohen T

Research Square, 13 Jan 2023

model of sporadic **TDP-43 proteinopathy** based on the principle that disease-associated **TDP-43** acetylation at... a model in which **TDP-43** acetylation drives neuronal dysfunction and **cognitive decline** through aberrant

Cited by: 0 articles | PPR: PPR596760

+ Add to export list

A Preprint v3

RNA-binding deficient TDP-43 drives cognitive decline in a mouse model of TDP-43 proteinopathy

Annotations (33)

Get citation

Claim to ORCID

Necarsulmer J, Simon J, Evangelista B, Chen Y, Tian X, Nafees S, Gonzalez AM, Wang P, Ajit D, Nikolova V, Harper K, Ezzell J, Beltran A, Moy S, Cohen T

Author information

Preprint from **Research Square**, 13 Jan 2023

DOI: 10.21203/rs.3.rs-2200020/v3 PPR: PPR596760

B Preprint v3

C This article is a preprint. It may not have been peer reviewed.

D A preprint is a complete scientific manuscript that an author uploads on a public server for free viewing. Initially it is posted without peer review, but may acquire feedback or reviews as a preprint, and may eventually be published in a peer-reviewed journal. The posting of preprints on public servers allows almost immediate dissemination and scientific feedback early in the 'publication' process.

E Preprint version history

- Version 3 [13 Jan 2023]
- Version 2 [09 Jan 2023]
- Version 1 [01 Nov 2022]

Fig 4

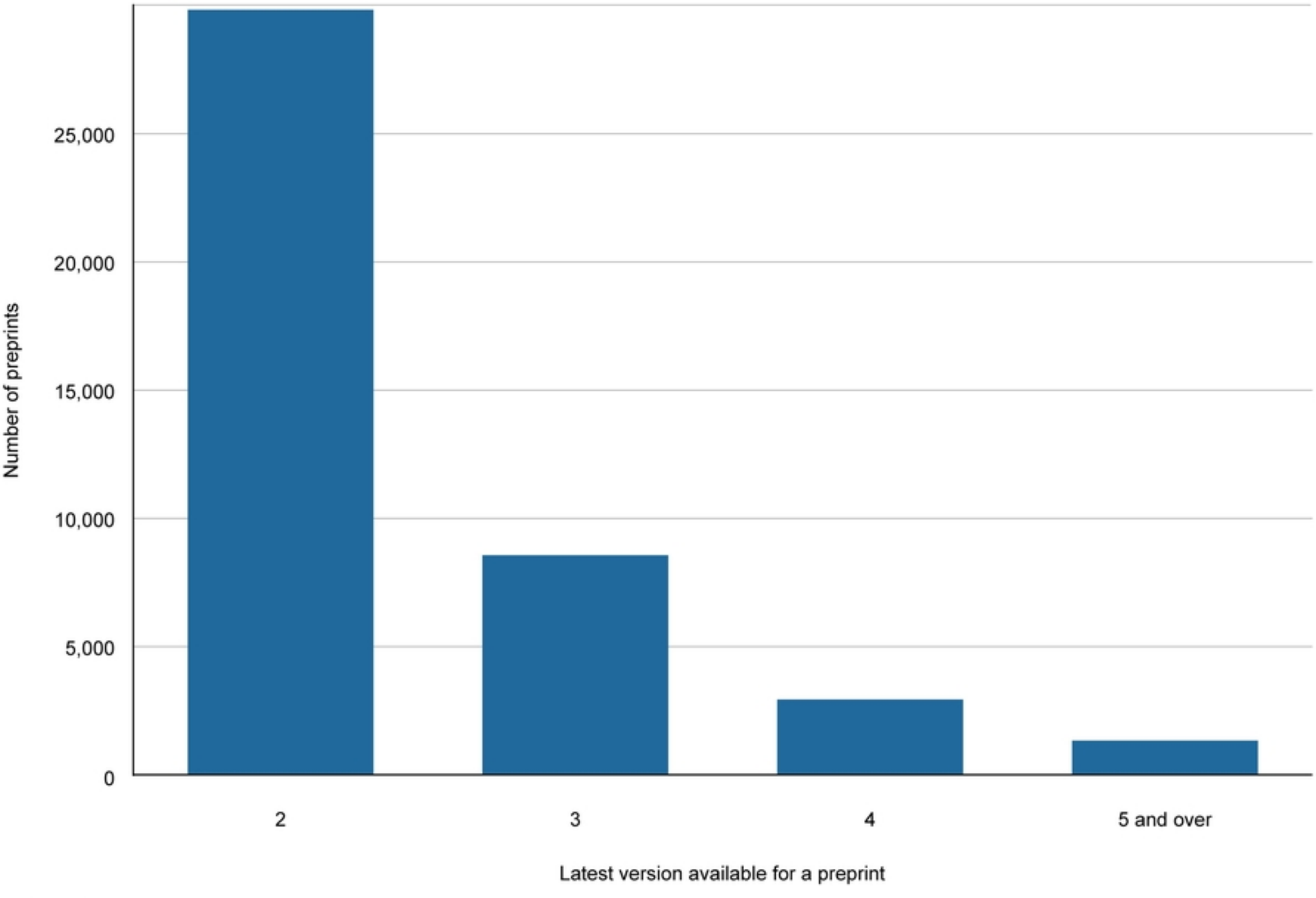
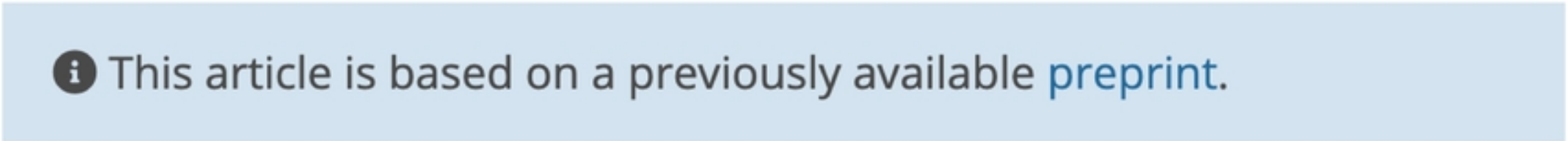

25,000
20,000
15,000
10,000
5,000
0
Number of preprints
2
3
4
5 and over
Latest version available for a preprint

Fig 5

A

This article is a preprint. A journal published article is available.

B


This article is based on a previously available preprint.

C

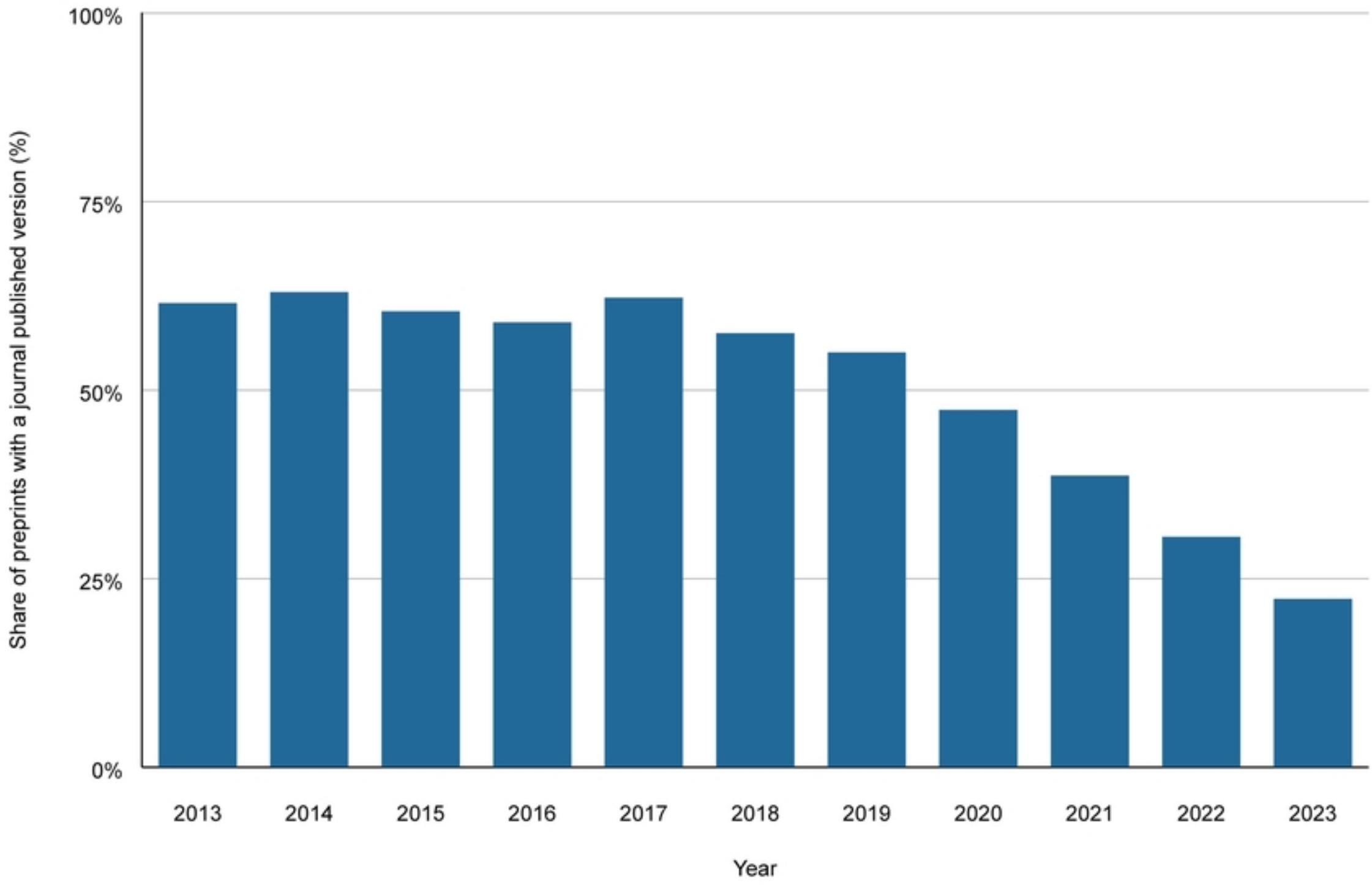

100%
75%
50%
25%
0%
Share of preprints with a journal published version (%)
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022
2023
Year

Fig 6

A

This article is a preprint. This preprint has been removed by the author(s).

B

This preprint has been withdrawn by the author(s). View this record on the preprint server for more information.

C

This article is a preprint. Version 3 of this preprint has been withdrawn by the author(s).

Fig 7

A

Abstract

Figures (5)

Free full text ▸

B

Citations & impact

Data

Reviews & evaluations

Funding

Non-neuronal expression of SARS-CoV-2 entry genes in the olfactory system suggests mechanisms underlying COVID-19-associated anosmia

Brann DH [1], Tsukahara T [1], Weinreb C [1], Lipovsek M [2], Van den Berge K [3], Gong B [4], Chance R [5], Macaulay IC [6], Chou H [5], Fletcher R [5], Das D [5], Street K [7], de Bezieux HR [4], Choi Y [8], Risso D [9], Dudoit S [3,4], Purdom E [4], Mill JS [10], Hachem RA [11], Matsunami H [12] ... [Show all 25] ... Datta SR [1]

Annotations

In full text (711)

C

Get citation

Open PDF

Claim to ORCID

Fig 8

A

⚠ This article is a preprint. Reviews & evaluations are available.

B

⚠ This article is undergoing peer review. Please access the article on the publisher's site to see the current peer review status.

C

⚠ This article is a preprint. Reviews & evaluations are available for a version(s) in the Version history below.

Preprint version history

- Version 2 [11 Oct 2021] , Reviews & evaluations available
- Version 1 [22 Sep 2021]

D

Reviews & evaluations

Reviews, recommendations & commentary from expert sources.

Rapid Reviews Infectious Diseases

26 Apr 2021
Review article
Ying-Rong Du, Hai-yan Fu, Ariel Berlinski

http://dx.doi.org/10.1162/2e3983f5.3790e6f9

26 Apr 2021
Review article
Ying-Rong Du, Hai-yan Fu

http://dx.doi.org/10.1162/2e3983f5.f99d4e65

26 Apr 2021
Review article
Ariel Berlinski

http://dx.doi.org/10.1162/2e3983f5.182368a8

MIT Press - Journals

26 Apr 2021
Reviews of "Inhaled budesonide in the treatment of early COVID-19 illness: a randomised controlled trial"
Hai-yan Fu, Ying-Rong Du, Ariel Berlinski

http://dx.doi.org/10.1162/2e3983f5.3790e6f9

Fig 9