



SOFTWARE TOOL ARTICLE

REVISED search.bioPreprint: a discovery tool for cutting edge, preprint biomedical research articles [version 2; peer review: 2 approved]

Carrie L. Iwema, John LaDue, Angela Zack, Ansuman Chattopadhyay

Health Sciences Library System, University of Pittsburgh, Pittsburgh, USA

v2 First published: 16 Jun 2016, 5:1396
<https://doi.org/10.12688/f1000research.8798.1>
 Latest published: 20 Jul 2016, 5:1396
<https://doi.org/10.12688/f1000research.8798.2>

Abstract

The time it takes for a completed manuscript to be published traditionally can be extremely lengthy. Article publication delay, which occurs in part due to constraints associated with peer review, can prevent the timely dissemination of critical and actionable data associated with new information on rare diseases or developing health concerns such as Zika virus. Preprint servers are open access online repositories housing preprint research articles that enable authors (1) to make their research immediately and freely available and (2) to receive commentary and peer review prior to journal submission. There is a growing movement of preprint advocates aiming to change the current journal publication and peer review system, proposing that preprints catalyze biomedical discovery, support career advancement, and improve scientific communication. While the number of articles submitted to and hosted by preprint servers are gradually increasing, there has been no simple way to identify biomedical research published in a preprint format, as they are not typically indexed and are only discoverable by directly searching the specific preprint server websites. To address this issue, we created a search engine that quickly compiles preprints from disparate host repositories and provides a one-stop search solution. Additionally, we developed a web application that bolsters the discovery of preprints by enabling each and every word or phrase appearing on any web site to be integrated with articles from preprint servers. This tool, search.bioPreprint, is publicly available at <http://www.hsls.pitt.edu/resources/preprint>.

Keywords

bookmarklet , grey literature , open science , peer review , pre-print , pre-publish , search engine , server

Open Peer Review

Approval Status

	1	2
version 2 (revision) 20 Jul 2016	 view	
version 1 16 Jun 2016	 view	 view

1. **Prachee Avasthi**, University of Kansas
 Medical Center, Kansas City, USA

2. **Cynthia Wolberger**, Johns Hopkins
 University School of Medicine, Baltimore, USA

Any reports and responses or comments on the article can be found at the end of the article.

Corresponding author: Ansuman Chattopadhyay (ansuman@pitt.edu)

Competing interests: No competing interests were disclosed.

Grant information: The author(s) declared that no grants were involved in supporting this work.

Copyright: © 2016 Iwema CL *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Iwema CL, LaDue J, Zack A and Chattopadhyay A. **search.bioPreprint: a discovery tool for cutting edge, preprint biomedical research articles [version 2; peer review: 2 approved]** F1000Research 2016, 5:1396 <https://doi.org/10.12688/f1000research.8798.2>

First published: 16 Jun 2016, 5:1396 <https://doi.org/10.12688/f1000research.8798.1>

REVISED Amendments from Version 1

We sincerely thank reviewers for their constructive comments and thoughtful suggestions. We incorporated several of their suggestions, as reflected in the Conclusions and Limitations sections in the revised version. Specifically,

We added a few sentences on using Google Scholar to search for preprints under Conclusions: “Google Scholar (GS), a popular scholarly literature search engine that provides cross-discipline search functionality, does not include preprint articles as a filter option. Hence, many avid GS users try a workaround by including preprint with the query term, (E.g., “asthma preprint” or “CRISPR preprint”) with the assumption of retrieving only preprint articles fetched from major preprint servers. In contrast, the GS search results in a mixed population of articles comprising both actual preprints and peer-reviewed published articles in which the term “preprint” appears somewhere in the full text of the article.”

We mentioned the need for reordering the search results by date by adding a new paragraph at the end of the Limitations section: “Currently, the search.bioPreprint default search results are ordered by relevance without any option to re-sort by date. The authors are aware of the pressing need for this added feature and if possible will incorporate it into the next version of the search tool.”

We also added a sentence under Conclusions: “Referees during the grant or journal article review process might also find this bookmarklet useful as it quickly retrieves pre-published articles via the cross-platform preprint search.”

See referee reports

Introduction

Preprint servers are online repositories that manage access to manuscripts that have not yet been peer-reviewed or formally published in a traditional manner. Preprint manuscripts are not copyedited, but they do undergo a basic screening process to check against plagiarism, offensiveness, and non-scientific content. Authors may make revisions at any point, but all versions remain available online. It should be noted that the term “preprint” in this context refers to manuscripts posted by the authors themselves onto specific online servers, not articles made available online by publishers a few weeks ahead of traditional publication.

Preprint articles can be more difficult to discover than those published traditionally, as they are not currently indexed in Medline and therefore do not appear in PubMed search results. This suggests that many timely and relevant research reports potentially fall through the cracks, as the time it takes to traditionally publish a biomedical manuscript can take anywhere from a few months to a few years. This lengthy process is seen by researchers to be a hindrance to scientific advancement. In response, there is a developing movement of preprint advocates who propose that preprints play a role in “catalyzing scientific discovery, facilitating career advancement, and improving the culture of communication within the biology community”¹. Preprint servers “enable authors to make their findings immediately available to the scientific community and receive feedback on draft manuscripts before they are submitted to journals”².

The history, rationale, and controversy surrounding preprint servers and the pace of the current publication process has been well

addressed in other manuscripts^{3–14}, news items^{15–22}, and blogs or white papers^{23–32}. We do not intend to duplicate this information here, but suggest exploration of our reference list for an overview of the current state of the topic.

Preprint server examples

There are currently only a small number of preprint servers catering to biological and biomedical research manuscripts.

- **arXiv** is a venerable preprint server covering physics, mathematics, computer science, nonlinear sciences, statistics, and quantitative biology since 1991. arXiv is funded by Cornell University Library, the Simons Foundation, and many member institutions.
- **bioRxiv**, operated by Cold Spring Harbor Laboratory, covers new, confirmatory, and contradictory results in research ranging from animal behavior and cognition to clinical trials, neuroscience to zoology.
- **F1000Research**, a member of the Science Navigation Group, provides an open science platform for the immediate publication of scientific communication. Posters and slides receive a digital object identifier and are instantly citable. Articles with associated source data are published within a week and made available for open peer review and user commenting. Articles that pass peer review are then indexed in PubMed, Scopus, and Google Scholar. It should be noted that *F1000Research* is not technically a preprint server, but is included here because it does provide access to articles prior to and during the peer review process. See the Limitations section for details.
- **PeerJ Preprints** covers biological, medical, life, and computer sciences. Their aim is to reduce publishing costs while still efficiently publishing innovative research, with an emphasis on not yet peer-reviewed articles, abstracts, or posters. Submissions are free, can be a draft, incomplete, or final version, and are typically online within a day after editorial approval.

Our intention is to present a resource that facilitates the quick and easy identification and access of scientific content located on preprint servers. **The Health Sciences Library System at the University of Pittsburgh** (HSLs) developed a tool to help researchers to quickly search preprint databases and discover cutting edge, yet-to-be published or reviewed biomedical research articles, **search.bioPreprint** (Figure 1). This search engine encompasses a federated search of arXiv, bioRxiv, *F1000Research*, and PeerJ Preprints. For ease of reading we will continue to refer to all sources of preprint articles as “preprint servers,” including the open science publishing platform *F1000Research*. We chose to publish this article in *F1000Research* and bioRxiv in order to support the preprint movement and to elicit feedback on usage of the tool, which will be updated as needed.

Implementation

Search engine

search.bioPreprint was created using the proprietary software **IBM Watson Explorer**, formerly Vivisimo Velocity, version 8.0-2 (IBM Corp, Armonk, New York, USA) to generate a meta search engine that compiles search term results from a pre-selected list of multiple sources into a single list ordered by the relevance of matching query terms. The results can then be further filtered by

search.bioPreprint

Comprehensively search preprint databases to discover cutting edge, yet-to-be published or reviewed biomedical research articles.

[About search.bioPreprint](#)

Search Examples: keyword (CRISPR); phrase ("Zika virus")
Use quotation marks for searches with 2 words or more.

Resources searched: [arXiv](#) (Quantitative Biology), [bioRxiv](#), [F1000Research](#), [PeerJ Preprints](#)

Install the [bioPreprint bookmarklet](#) to integrate seamlessly any text from any web page with the bioPreprint search engine!



[Contact the administrator with comments or questions](#)

Figure 1. Website homepage for search.bioPreprint.

Source (e.g., the preprint servers of origin) or by Topic (e.g., microcephaly for a Zika virus search). The Topic search is accomplished via clustering, meaning the search results are organized on the fly by similarity in subject matter. Additionally, a “remix” link displayed next to the clustered topics reveals new secondary topics. This is done by clustering the same search results again, but explicitly ignoring the topics that were used in the initial clustering process.

The Health Sciences Library System at the University of Pittsburgh has repeatedly utilized IBM Watson Explorer software to develop, implement, and maintain several federated search engines focused on a variety of topics. These include: [search.HSL.OBRC](#) –a portal for discovering bioinformatics databases and software via the Online Bioinformatics Resource Collection³³, [Clinical Focus](#) –a portal providing quick access to high-quality clinical information³⁴, and [Clinical eCompanion](#) –a portal with information for primary care³⁵. Similarly, the U.S. National Library of Medicine (NLM) [utilized the same software](#) to create search engines for [MedlinePlus](#), [MedlinePlus en Español](#), and the [NLM library website](#).

The search engine was created following the software manufacturer’s protocol. Briefly, the search url and parameters are entered for each site, then the results are selected based on the XPath of the results within the HTML page. Finally, each individual source is bundled into a single source to provide one search for multiple sites. A maximum of 200 total results are returned based on the licensing agreement with IBM; this also contributes to a short wait for return of results. The selected sources for retrieving

preprint articles using search.bioPreprint are: (1) the quantitative biology section of [arXiv.org](#), (2) [bioRxiv](#), (3) [F1000Research](#) and (4) [PeerJ Preprints](#).

How-to-use

As an example, typing a single-word query term, such as CRISPR, into the search box results in **ninety-one preprint articles** culled from the aforementioned preprint servers (Figure 2, searched on 2 May 2016). Clicking on an article title redirects to that article at its original source. Search results may be narrowed by Topic or Source using the filters on the left side of the page. Using the CRISPR example, the ninety-one search results are grouped into shared Topics: fourteen articles on “Bacterial,” twelve articles on “Protein,” six articles on “Genome engineering,” etc. Expanding individual topics reveals a list of subtopics: clicking on the topic “Protein” redistributes the twelve articles into subtopics, including “CRISPR-Cas9,” “Image, Palindromic Repeat,” “Mutants, Generated,” etc. Clicking on a topic or subtopic reconfigures the search results to limit to these filtered articles.

Clicking on the “remix” button appearing next to “Top 91 results” regroups the original search results into additional topics such as “Cells,” “Advances,” “Drosophila,” etc that are not present in the first results iteration (Figure 3). This provides another opportunity to discover pertinent preprint articles, especially if a large number of results is returned.

The search results may also be filtered by Source. Selecting this will change the default display of topic-focused clusters to articles



CRISPR

Search Examples: keyword ([CRISPR](#)); phrase ("[Zika virus](#)")
Use quotation marks for searches with 2 words or more.

Resources searched: [arXiv](#) (Quantitative Biology), [bioRxiv](#), [F1000Research](#), [PeerJ Preprints](#)

Install the [bioPreprint bookmarklet](#) to integrate seamlessly any text from any web page with the bioPreprint search engine!

Narrow your search by:

[Topic](#) [Source](#)

Top 91 Results remix

- + Bacterial (14)
- Protein (12)
 - Recent advances (3)
 - CRISPR-Cas9 (3)
 - Image, Palindromic Repeat (2)
 - Mutants, Generated (2)
 - Expression, Cas9-Mediated (2)
 - Other Topics (1)
- + Genome engineering (6)
- + Modeling (5)
- + RNA And Crispr (5)
- + Clustered regularly interspaced short palindromic repeats (5)
 - Crispr-Cas9 In Gene (2)
 - CRISPR screen (3)
 - Chromosomal (3)
 - Guide RNAs (2)

[more](#) | [all](#)

Search within clusters

1. [Recent advances in large-scale protein interactome mapping](#)
affinity tagging of endogenous proteins using the powerful CRISPR/Cas9 (clustered regularly interspaced ...
F1000Research
2. [SwissPalm: Protein Palmitoylation database](#)
using the CRISPR-CAS9 system) when compared to control cells (Figure 8D). Conclusion As the first ...
F1000Research
3. [Strong morphological defects in conditional Arabidopsis abp1 knock-down mutants generated in absence of functional...](#)
the CRISPR-induced small deletion does not necessarily decrease transcript levels - is probably cause ...
F1000Research
4. [Comparative assessment of fluorescent proteins for in vivo imaging in an animal model system](#)
...Goldstein bobg@email.unc.edu Characters: 19,566 2 Abbreviations: FP – fluorescent protein GFP – green fluorescent protein mNG – monon
bioRxiv
5. [Drosophila CLAMP is an essential protein with sex-specific roles in males and females](#)
...studies on the in vivo function of CLAMP remained difficult because the location of the gene in pericentric heterochromatin made it challeng
bioRxiv
6. [Heterozygous gene truncation delineates the human haploinsufficient genome](#)
...fewer paralogs, were more likely to be part of protein complexes and were more connected in protein-protein interaction networks than the re
bioRxiv
7. [Dilp8 requires the neuronal relaxin receptor Lgr3 to couple growth to developmental timing](#)
...couples growth to developmental timing 6 Lgr3 is expressed in the CNS To gain insight into the tissue and cellular expression pattern of Lgr3
bioRxiv

Figure 2. Search results page with query term CRISPR. At left is the default view by Topic. (2 May 2016).

Narrow your search by:

[Topic](#) [Source](#)

Top 91 Results remix

- + Cells (24)
- + Advances (7)
- + Populations (6)
- + Cas System (5)
- + Libraries (5)
- + Drosophila (5)
- + Spacer (5)
- + Viral (5)
 - Immunity (4)
 - Gene editing (4)

[more](#) | [all](#)

organized by Source, which in the current iteration is one of the four preprint servers searched by this tool: nineteen from *F1000Research*, two from PeerJ Preprints, six from arXiv, and sixty-five from bioRxiv (Figure 4).

Quotation marks are recommended for searches with exact phrases, e.g., Zika virus. The necessity of this was discovered after examining the search parameters of the various preprint servers. As one of the preprint servers by default joins words in a multi-word query with the Boolean operator “OR” then a search for a phrase such as zika virus produces multiple articles where the only matching term is virus. Using quotation marks for a search of more than one word mitigates this problem and considerably improves the quality of results. A search for “zika virus” thus produces **seventy-nine articles** that are typically filtered into “Zika virus infection,” “Microcephaly,” “Discovery,” “Dengue Virus,” etc (searched on 2 May 2016).

Figure 3. Topics change after selecting “remix.” (2 May 2016).

Narrow your search by:

[Topic Source](#)

Top 92 Results

- [F1000Research](#) (19)
- [PeerJ Preprints](#) (2)
- [arXiv](#) (6)
- [bioRxiv](#) (65)

Figure 4. Results view by preprint server Source. (2 May 2016).

The “Search within clusters” box allows for searching within the search results, and can be used to identify specific articles within the cohort of Zika virus preprints that are not immediately apparent from topical clustering. Entering vaccine in the search box highlights the topics and subtopics containing articles bearing the word vaccine: under “Zika virus infection” is “Preventing Zika Virus Infection;” under Dengue Virus is “Antibodies, Vaccine” and “Community, Vector.” Selection of highlighted topics or subtopics reconfigures the results to limit to vaccine-related Zika virus preprints (Figure 5).

bioPreprint-bookmarklet

A bookmarklet is a special type of web browser widget containing an embedded software command that extends the application of the browser by adding a one-click function as a bookmark. We created a [bioPreprint-bookmarklet](#) using JavaScript in order to seamlessly integrate a search for any word or phrase from any web page with the information stored in preprint servers. After dragging/dropping the bioPreprint-bookmarklet into any web browser, the next step is to highlight a word or phrase of interest then click the bookmarklet. This will result in a pop-up window displaying preprint articles containing the text of interest (Figure 6).

All web browsers that support JavaScript (Google Chrome, Mozilla FireFox, Internet Explorer, Apple Safari, Opera) are compatible with the bookmarklet. In case the favorites/bookmark bar is not visible we provide [instructions](#) for displaying it on commonly used browsers. A [video](#) describing how to install the bookmarklet in a web browser is also available.

Use cases

Scenario 1

Imagine a researcher is searching PubMed for articles on “RNA-seq quantification” and comes across a paper recently

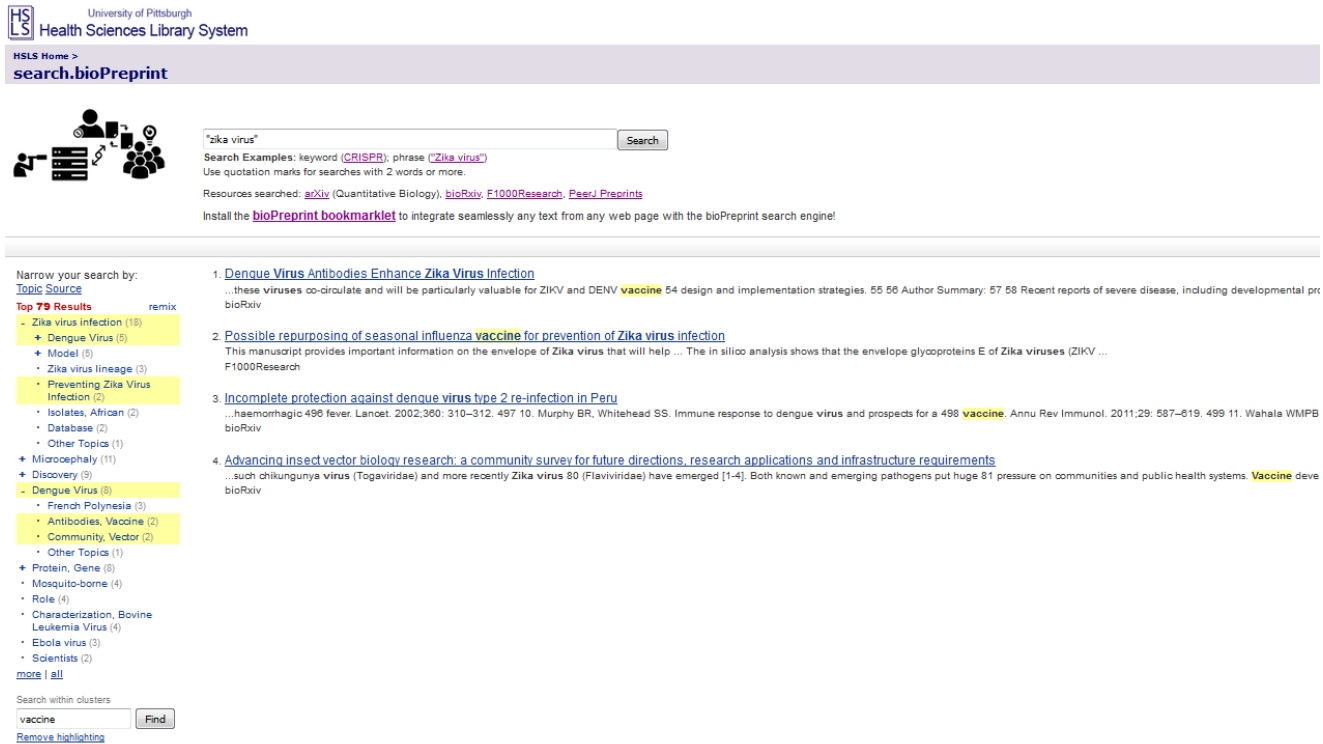


Figure 5. Results for Zika virus using quotation marks and the “Search within clusters” feature. (2 May 2016).

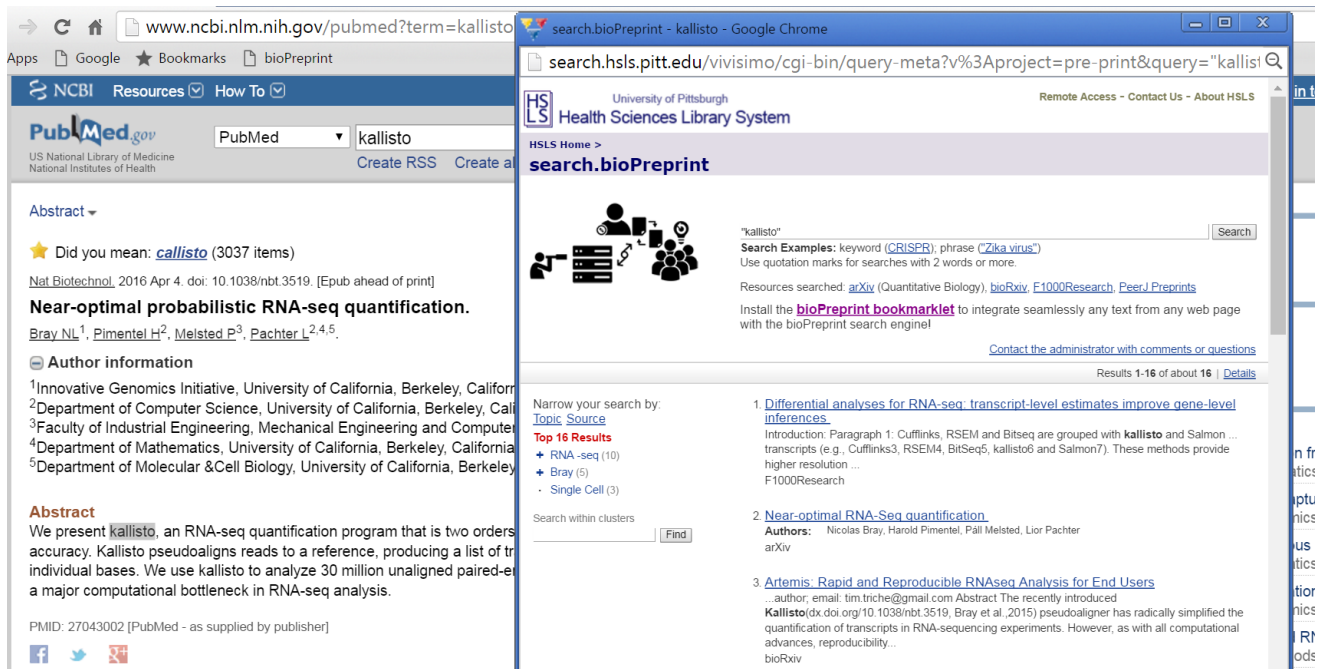


Figure 6. Using the bioPreprint-bookmarklet. (2 May 2016).

published in Nature Biotechnology, “Near-optimal probabilistic RNA-seq quantification”³⁶. This paper introduces a new software program, Kallisto, that analyzes RNA-seq data by two orders of magnitude faster than previously used software. This is notable as it removes the computational bottleneck for RNA-seq data analysis. After reading about this new software, the researcher decides to check whether it has been widely adopted by perusing the published literature.

A search in PubMed with the search term “Kallisto” results in only the original article (searched on 2 May, 2016). This is well within expectations, considering the recent publication date of the article, 4 April 2016. There has not been enough time for researchers to know about the software, let alone write papers citing it.

To continue to try and gauge the usage of Kallisto in RNA-seq data analysis, the researcher might take an alternative approach: instead of searching PubMed, try searching for preprint articles. This can be achieved with a single click of the bioPreprint-bookmarklet once it is installed in the researcher’s web browser. Upon viewing the article abstract on the PubMed search results page, highlighting the word “Kallisto,” and clicking the bioPreprint-bookmarklet, a pop-up appears with the search.bioPreprint search results: sixteen preprint articles, two from arXiv, thirteen from bioRxiv, and one from *F1000Research* (searched on 2 May, 2016). Interestingly, the second article on the results page is the preprint version of the Nature Biotechnology paper on Kallisto software, submitted to the arXiv preprint server (Figure 6). The authors submitted their preprint on 11 May 2015, almost one year before its publication in Nature Biotechnology, with concomitant indexing by PubMed³⁷.

It is worth noting that since the availability of the Kallisto paper as a preprint, fifteen preprint articles have cited the use of Kallisto software^{38–51}, searched on 2 May, 2016). These articles cover numerous topics, including development of new software, single cell RNA-seq analysis, and quantification of the relative abundance of transcripts in various experimental settings.

Scenario 2

A student gathering information from the internet about the regulation of gene expression happens upon the GTEx Project Community Scientific Meeting website. GTEx stands for the Genotype-Tissue Expression project (GTEx), which aims to develop an atlas of human gene expression and its regulation across various tissue types. Intrigued by the scope of this project, the student is curious to know how GTEx project data have been utilized in research.

The bioPreprint search engine and bookmarklet can quickly satisfy the student’s curiosity by providing easy access to GTEx-related articles hosted by various preprint servers that may or may not be published “in print” yet. This process is simple, unique, and the student doesn’t even need to leave the current web page to go on a literature hunt. Rather, all GTEx-related articles will appear in a new window with only two clicks, the first highlighting the word GTEx and the second on the previously-installed bioPreprint-bookmarklet. The result is sixty-seven articles showcasing the use of GTEx data in a variety of research topics including “Genome Wide Association Studies,” “Allele, Specific expression,” “Expression Quantitative Trait Loci,” etc (searched on 2 May 2016).

Conclusions

These use cases emphasize the power of the bioPreprint search engine and associated bookmarklet in delivering scientific research articles that are not only hard-to-find and yet-to-be traditionally published, but also on demand at the point of reading. And the “point of reading” can be anything on the web: journal articles, news items, blogs, PubMed/Google Scholar search results, etc.

Until the creation of search.bioPreprint there has been no simple and efficient way to identify biomedical research published in a preprint format, as they are not typically indexed and are primarily discoverable by directly searching the preprint server websites (articles that pass peer review in *F1000Research* are the exception). [Google Scholar](#) (GS), a popular scholarly literature search engine that provides cross-discipline search functionality, does not include preprint articles as a filter option. Hence, many avid GS users try a workaround by including preprint with the query term, (E.g., “asthma preprint” or “CRISPR preprint”) with the assumption of retrieving only preprint articles fetched from major preprint servers. In contrast, the GS search results in a mixed population of articles comprising both actual preprints and peer-reviewed published articles in which the term “preprint” appears somewhere in the full text of the article.

During the final stages of manuscript preparation an online database aiming to index preprint articles was launched, [PrePubMed](#), which despite appearances is not an official resource from the National Library of Medicine (NLM), the National Center for Biotechnology Information (NCBI), or PubMed. We want to acknowledge this new resource, but emphasize that search.bioPreprint offers full text searching where available (currently, bioRxiv and *F1000Research*, and arXiv in the future) as well as topical and source-based clustering of results. In addition, our tool has been available since mid-February 2016, around the same time as the [ASAPbio meeting](#), where it was mentioned during discussions.

The underlying technology upon which search.bioPreprint was built is flexible enough to integrate additional resources into the search engine. As new preprint servers are introduced, search.bioPreprint will incorporate them and continue to provide a simple solution for finding preprint articles. We welcome feedback that introduces new preprint resources and addresses usability concerns.

The bioPreprint-bookmarklet enables each and every word or phrase appearing on any website to be integrated with information in articles stored in preprint servers. The on-demand delivery of preprint articles at the point of reading enables researchers to discover brand new pre-published articles quickly and be updated with cutting edge, yet-to-be-reviewed information that is challenging to discover by traditional literature searching methods. Referees during the grant or journal article review process might also find this bookmarklet useful as it quickly retrieves pre-published articles via the cross-platform preprint search.

Our intention is that the combined use of the aforementioned tools helps to fulfill the unmet need of the scientific community for

immediate dissemination of research outcomes, ultimately resulting in improved scientific communication and far-ranging insights and innovations.

Limitations

While arXiv, bioRxiv, and PeerJPreprints are considered to be preprint servers, *F1000Research* belongs to a separate class. It offers a unique publishing platform in which a transparent peer review process is integrated into the article publication practice and thus holds three categories of articles based on peer review status: (1) recently submitted and awaiting peer review, (2) passed peer review, and (3) not passed by peer reviewers. Only articles that pass the peer review process are indexed in literature databases such as PubMed. *F1000Research* permanently hosts all articles irrespective of peer review status. Therefore, it represents a blended system of preprint server and traditional online journal. Search.bioPreprint does not separate these three types of *F1000Research* articles and therefore returns both non-peer reviewed and reviewed articles together in the search results. Nevertheless, the peer review status is easily visible when searchers are directed to the *F1000Research* site from the search.bioPreprint search results. As *F1000Research* hosts many articles whose peer review status (before passing peer review) could be considered the equivalent of preprints, we decided to include this as a source of preprint articles. Users should note a key difference, however, as all articles in *F1000Research* are committed to formal peer review and should therefore not be submitted to any additional journals.

The quality of the search results generated by the bioPreprint search engine is confined by the search parameters of the individual preprint servers. If the preprint servers alter their search algorithms, a concomitant adjustment of underlying codes used by the bioPreprint search engine is often required. Unfortunately, this can be done without any public notification and is only discoverable upon a thorough analysis of bioPreprint search results. The University of Pittsburgh Health Sciences Library System has a quality check team involving two librarians to ensure the accuracy of search.bioPreprint results. The team routinely compares the search results produced by several preset query terms with the previous results and reports any discrepancies to the development team.

The average time taken to display search results is not always optimal. The speed of the search.bioPreprint results return stems from multiple factors: individual preprint servers’ searching speed, efficiency of the IBM Watson Explorer software, and computational power of the server hosting the bioPreprint search engine. While some contributing factors are outside of our control, efforts will be undertaken to speed up the search process by continually upgrading the power of the host server.

Currently, the search.bioPreprint default search results are ordered by relevance without any option to re-sort by date. The authors are aware of the pressing need for this added feature and if possible will incorporate it into the next version of the search tool.

Software availability

search.bioPreprint is freely accessible at <http://www.hsls.pitt.edu/resources/preprint>. The preprint search engine was created using the software, IBM Watson Explorer, formerly known as Vivisimo Velocity. IBM Watson Explorer is a proprietary software, hence, its source code is not available.

The bioPreprint-bookmarklet is freely available at <http://hsls.pitt.edu/biopreprint-infobooster>.

The JavaScript code embedded in the bookmarklet is:

```
“javascript:(function(){(function(t,u,w){t=''+(window.getSelection%3Fwindow.getSelection():document.getSelection%3Fdocument.getSelection():document.selection%3Fdocument.selection.createRange().text:');u=%3F'http://search.hsls.pitt.edu/vivisimo/cgi-bin/query-meta%3Fv%253Aproject=preprint%26query=%2522'+encodeURIComponent(t)+'%2522':";w=window.open(u,'_blank','height=750,width=700,scrollbars=1');w.focus()%;26%26 w.focus();if(!t){w.document.write('<html><head><title></title></head><body style="padding:1em;font-family:Helvetica,Arial"><br /><p>First%2C highlight a word or a group of words from any website that you are browsing (journal article%2C PubMed search result%2C news article%2C blog%2C etc.)%2C and then click on this bookmarklet to retrieve cutting edge%2C yet-to-be published or reviewed biomedical research articles related to your selected word(s).</a><p>Check the <a href="http://media.hsls.pitt.edu/media/BioPreprint_ac0316.mp4">How to Video</a>for instruction.</p><br/><p></p><script>var q=document.getElementById("q"),v=q.
```

```
value;q.focus();q.value="";q.value=v;</script></body></html>');w.document.close();})();})();”
```

Author contributions

AC conceived the concept and wrote the Implementation section. CI created the search.bioPreprint logo, assisted with concept refinement and design, and wrote all documentation, including preparation of the initial drafts of the manuscript. AC and CI created the figures. JL developed the search engine. AZ created the bookmarklet and webpage. All authors were involved in revision of the draft manuscript and have agreed to the final content.

Competing interests

No competing interests were disclosed.

Grant information

The author(s) declared that no grants were involved in supporting this work.

Acknowledgements

The authors wish to gratefully acknowledge the following individuals for their help with various aspects of the creation of search.bioPreprint and manuscript preparation: Peter Coles for writing a [blog](#) on the insightful use of bookmarklets, Julia Dahm for the creation of the video describing how to install the bookmarklet, Melissa Ratajeski for providing helpful comments on the manuscript, Nancy Tannery for providing helpful comments on the manuscript and offering general support for this project, and Fran Yarger for offering general support for this project.

References

- ASAPbio: **Accelerating Science and Publication in Biology**. [cited 2016 Mar 29]. [Reference Source](#)
- bioRxiv. [cited 2016 Mar 29]. [Reference Source](#)
- Desjardins-Proulx P, White EP, Adamson JJ, *et al.*: **The case for open preprints in biology**. *PLoS Biol.* 2013; **11**(5): e1001563. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Fisher D, Parisi N: **Social influence and peer review: Why traditional peer review is no longer adapted, and how it should evolve**. *EMBO Rep.* 2015; **16**(12): 1588–1591. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Ford E: **Open peer review at four STEM journals: an observational overview [version 2; referees: 2 approved, 2 approved with reservations]**. *F1000Res.* 2015; **4**: 6. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Hu C, Zhang Y, Chen G: **Exploring a New Model for Preprint Server: A Case Study of CSPO**. *J Acad Libr.* 2010; **36**(3): 257–262. [Publisher Full Text](#)
- Hutchins BI, Yuan X, Anderson JM, *et al.*: **Relative Citation Ratio (RCR): A new metric that uses citation rates to measure influence at the article level**. *bioRxiv.* 2015; 029629. [Publisher Full Text](#)
- Lauer MS, Krumholz HM, Topol EJ: **Time for a prepublication culture in clinical research?** *Lancet.* 2015; **386**(10012): 2447–2449. [PubMed Abstract](#) | [Publisher Full Text](#)
- McDowell G: **Junior biomedical scientists and preprints [version 1; referees: 1 approved with reservations]**. *F1000Res.* 2016; **5**: 294. [Publisher Full Text](#)
- Powell K: **Does it take too long to publish research?** *Nature.* 2016; **530**(7589): 148–151. [PubMed Abstract](#) | [Publisher Full Text](#)
- Tomauiolo NG, Packer JG: **Preprint Servers: Pushing the Envelope of Electronic Scholarly Publishing**. *Search Medford N J.* 2000; **8**(9). [Reference Source](#)
- Tracz V, Lawrence R: **Towards an open science publishing platform [version 1; referees: 2 approved]**. *F1000Res.* 2016; **5**: 130. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Vale RD: **Accelerating scientific publication in biology**. *Proc Natl Acad Sci U S A.* 2015; **112**(44): 13439–13446. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Warne V: **Rewarding reviewers - sense or sensibility? A Wiley study explained**. *Learned Publishing.* 2016; **29**(1): 41–50. [Publisher Full Text](#)
- Callaway E: **Preprints come to life**. *Nature.* 2013; **503**(7475): 180. [PubMed Abstract](#) | [Publisher Full Text](#)
- Callaway E: **Geneticists eye the potential of arXiv**. *Nature.* 2012; **488**(7409): 19. [PubMed Abstract](#) | [Publisher Full Text](#)
- Callaway E, Powell K: **Biologists urged to hug a preprint**. *Nature.* 2016; **530**(7590): 265. [PubMed Abstract](#) | [Publisher Full Text](#)

18. Gibney E: **Open journals that piggyback on arXiv gather momentum.** *Nature*. 2016; **530**(7588): 117–118.
[PubMed Abstract](#) | [Publisher Full Text](#)
19. Harmon A: **Handful of Biologists Went Rogue and Published Directly to Internet.** 2016. [cited 2016 Mar 28].
[Reference Source](#)
20. Kaiser J: **New Preprint Server Aims to Be Biologists' Answer to Physicists' arXiv.** *Science*. AAAS. 2013. [cited 2016 Mar 29].
[Reference Source](#)
21. Palmer KM: **A Rainbow Unicorn Wants to Transform Biology Publishing.** *WIRED*. 2016. [cited 2016 Mar 28].
[Reference Source](#)
22. **Taking the online medicine.** *The Economist*. 2016. [cited 2016 Mar 28].
[Reference Source](#)
23. Birney E: **10,000 Up.** 2015. [cited 2016 Mar 29].
[Reference Source](#)
24. Brown CT: **A good way to publish -- arXiv FTW.** 2012. [cited 2016 Mar 29].
[Reference Source](#)
25. Curry S: **Combining preprints and post-publication peer review: a new (big) deal?** *Reciprocal Space*. 2016. [cited 2016 Mar 29].
[Reference Source](#)
26. Himmelstein D: **The History of Publishing Delays.** 2016. [cited 2016 Mar 29].
[Reference Source](#)
27. Thomason A: **Is Scientific Publishing About to Be Disrupted? ASAPbio, Briefly Explained – The Ticker - Blogs - The Chronicle of Higher Education.** 2016. [cited 2016 Mar 28].
[Reference Source](#)
28. White E: **Which preprint server should I use?** *Jabberwocky Ecology*. The Weecology Blog on WordPress.com. 2014. [cited 2016 Mar 29].
[Reference Source](#)
29. Tracz V, Lawrence R: **The role of preprints in publishing.** *ASAPbio*. 2016. [cited 2016 Mar 29].
[Reference Source](#)
30. Smith R: **A better way to publish science-BMJ Blogs.** 2015. [cited 2016 Mar 30].
[Reference Source](#)
31. Eisen M, Vosshall LB: **Coupling Pre-Prints and Post-Publication Peer Review for Fast, Cheap, Fair, and Effective Science Publishing.** *ASAPbio*. 2016. [cited 2016 Mar 30].
[Reference Source](#)
32. Heard S: **Post-publication peer review and the problem of privilege.** *Scientist Sees Squirrel* on WordPress.com. 2015. [cited 2016 Mar 30].
[Reference Source](#)
33. Chen YB, Chattopadhyay A, Bergen P, *et al.*: **The Online Bioinformatics Resources Collection at the University of Pittsburgh Health Sciences Library System—a one-stop gateway to online bioinformatics databases and software tools.** *Nucleic Acids Res*. 2007; **35**(Database issue): D780–D785.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
34. Tannery NH, Epstein BA, Wessel CB, *et al.*: **Impact and user satisfaction of a clinical information portal embedded in an electronic health record.** *Perspect Health Inf Manag*. 2011; **8**: 1d.
[PubMed Abstract](#) | [Free Full Text](#)
35. Wessel C, LaDue J, Dahm J: **Clinical ECompanion: Development of a Point-of-Care Information Tool.** Toronto, Ontario, Canada: Medical Library Association; 2016.
[Reference Source](#)
36. Bray NL, Pimentel H, Melsted P, *et al.*: **Near-optimal probabilistic RNA-seq quantification.** *Nat Biotechnol*. 2016; **34**(5): 525–7.
[PubMed Abstract](#) | [Publisher Full Text](#)
37. Bray N, Pimentel H, Melsted P, *et al.*: **Near-optimal RNA-Seq quantification.** *arXiv.org*. 2015.
[Reference Source](#)
38. Arakawa K, Yoshida Y, Tomita M: **Genome sequencing of a single tardigrade *Hypsibius dujardini* individual.** *bioRxiv*. 2016.
[Publisher Full Text](#)
39. Gibilisco L, Zhou QI, Mahajan S, *et al.*: **The evolution of alternative splicing in *Drosophila*.** *bioRxiv*. 2016.
[Publisher Full Text](#)
40. Havens LA, MacManes MD: **Characterizing the Adult and Larval Transcriptome of the Multicolored Asian Lady Beetle, *Harmonia axyridis*.** *bioRxiv*. 2015.
[Publisher Full Text](#)
41. Hensman J, Papastamoulis P, Glaus P, *et al.*: **Fast and accurate approximate inference of transcript expression from RNA-seq data.** *arXiv.org*. 2015.
[Reference Source](#)
42. Kordonowy LL, MacManes MD: **Characterization of a Male Reproductive Transcriptome for *Peromyscus eremicus* (Cactus mouse).** *bioRxiv*. 2016.
[Publisher Full Text](#)
43. MacManes MD: **Establishing evidenced-based best practice for the de novo assembly and evaluation of transcriptomes from non-model organisms.** *bioRxiv*. 2016.
[Publisher Full Text](#)
44. Morgan AP, Holt JM, McMullan RC, *et al.*: **The evolutionary fates of a large segmental duplication in mouse.** *bioRxiv*. 2016.
[Publisher Full Text](#)
45. Ntranos V, Kamath GM, Zhang J, *et al.*: **Fast and accurate single-cell RNA-Seq analysis by clustering of transcript-compatibility counts.** *bioRxiv*. 2016.
[Publisher Full Text](#)
46. Pai AA, Baharian G, Sabourin AP, *et al.*: **Widespread shortening of 3' untranslated regions and increased exon inclusion are evolutionarily conserved features of innate immune responses to infection.** *bioRxiv*. 2016.
[Publisher Full Text](#)
47. Rzepiela AJ, Vina-Vilaseca A, Breda J, *et al.*: **Exploiting variability of single cells to uncover the *in vivo* hierarchy of miRNA targets.** *bioRxiv*. 2015.
[Publisher Full Text](#)
48. Patro R, Duggal G, Kingsford C: **Accurate, fast, and model-aware transcript expression quantification with Salmon.** *bioRxiv*. 2015.
[Publisher Full Text](#)
49. Soneson C, Matthes KL, Nowicka M, *et al.*: **Differential transcript usage from RNA-seq data: isoform pre-filtering improves performance of count-based methods.** *bioRxiv*. 2015.
[Publisher Full Text](#)
50. Soneson C, Love MI, Robinson MD: **Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences [version 2; referees: 2 approved].** *F1000Res*. 2016; **4**: 1521.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
51. Stubbington MJT, Lönnberg T, Proserpio V, *et al.*: **Simultaneously inferring T cell fate and clonality from single cell transcriptomes.** *bioRxiv*. 2015.
[Publisher Full Text](#)

Open Peer Review

Current Peer Review Status:  

Version 2

Reviewer Report 21 July 2016

<https://doi.org/10.5256/f1000research.9946.r15095>

© 2016 Avasthi P. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Prachee Avasthi

Department of Anatomy and Cell Biology, University of Kansas Medical Center, Kansas City, KS, USA

The authors have adequately addressed my comments. It is interesting to note that a Google Scholar search for "[search term] preprint" does not restrict results exclusively to preprints. While an advanced search in Google Scholar for articles published in only preprint servers would likely circumvent this, it is far more cumbersome to list all preprint servers in a Google Scholar advanced search (along with any new preprint servers that are developed) than to use search.Biopreprint (which will do this automatically). With the addition of a sort by date feature, this tool and associated bookmarklet will be an essential addition to the workflow of researchers.

Competing Interests: No competing interests were disclosed.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Version 1

Reviewer Report 06 July 2016

<https://doi.org/10.5256/f1000research.9471.r14408>

© 2016 Wolberger C. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Cynthia Wolberger

Department of Biophysics and Biophysical Chemistry, School of Medicine, Johns Hopkins University School of Medicine, Baltimore, MD, USA

This article contains a thorough description of a new search tool, search.bioPreprint, that can be used to search multiple preprint archives using keywords. In light of the existence of multiple preprint servers that can be used to post preprints in the biological sciences, the development of a tool that enables full-text searching of all current preprint archives (arXiv, bioRxiv, F1000Research, PeerJ Preprints) is a welcome one. The search capabilities of search.bioPreprint and the bookmarklet app are well-described.

Looking ahead, the authors will hopefully consider further improvements to the search site. Better documentation on the search site itself explaining how results are returned and ranked would be helpful. This reviewer quickly learned that the search returns approximate word match results, not just exact matches; this should be clarified on the web site. Moreover, approximate matches sometimes appear to be ranked more highly than exact matches; the reason for this should be examined and remedied, if possible. It would also be helpful to have options to rank results in other ways, in particular, by date.

Competing Interests: No competing interests were disclosed.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Author Response 13 Jul 2016

Ansuman Chattopadhyay, University of Pittsburgh, Pittsburgh, USA

We sincerely thank the reviewer for considering the manuscript as "Approved" and for the constructive review.

Reviewer Comment: "...the authors will hopefully consider further improvements to the search site. Better documentation on the search site itself explaining how results are returned and ranked would be helpful."

Author's Response:

We appreciate the suggestions and will take measures to improve documentation on the search site.

Reviewer Comment: "This reviewer quickly learned that the search returns approximate word match results, not just exact matches; this should be clarified on the web site. Moreover, approximate matches sometimes appear to be ranked more highly than exact matches; the reason for this should be examined and remedied, if possible."

Author's Response:

We agree that better documentation is needed and will add a thorough description of the search process on the search site. One of the limitations of this search engine is it completely depends on the search prowess of the individual sources and we do not have any control over that. If the query term is placed within quotation marks, it forces the search engines to apply an exact word matching algorithm thus mitigating the issue of approximate word matching.

Reviewer Comment: "It would also be helpful to have options to rank results in other ways, in particular, by date."

Author's Response:

We acknowledge the pressing need for an option to sort preprint articles by date as both reviewers mentioned it. We are actively working to bring this functionality into search.bioPreprint.

Competing Interests: No competing interests were disclosed.

Reviewer Report 04 July 2016

<https://doi.org/10.5256/f1000research.9471.r14404>

© 2016 Avasthi P. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Prachee Avasthi

Department of Anatomy and Cell Biology, University of Kansas Medical Center, Kansas City, KS, USA

search.bioPreprint is a useful tool for the scientific community and this article nicely outlines the key features and use cases for the service. A particular strength of the tool is the site maintenance and quality control by the University of Pittsburgh Health Sciences Library System. Also, the flexibility to incorporate new preprint servers due to continued site support is a benefit. The development of a bookmarklet for this purpose is novel, convenient and very useful.

I have a few comments for authors to address:

1. There is no mention in the abstract that preprints can be found through Google Scholar. While there is a temporary bug associated with Google Scholar's treatment of subsequently published preprints (documented in the following blog posts by Dr. Wilke), preprints are indeed searchable across platforms through Google Scholar. Please mention this and note any benefits of your tool. Google Scholar is mentioned on p2 but only in the context of indexing peer reviewed publications.

<http://serialmentor.com/blog/2014/11/1/the-google-scholar-preprint-bug>

<http://serialmentor.com/blog/2015/10/8/Google-Scholar-bug-redux>

2. The search results on search.bioPreprint are ordered by relevance with no option to re-sort. After some time searching using search.bioPreprint, the need to sort results by date became quickly evident. Likely, many users will be interested to know what new preprints are available across platforms since a previous search. A search term on Google Scholar

followed by “preprint” and one click on the “sort by date” link produces date-sorted preprints from multiple sources (biorxiv, arrive, peerj etc). Similarly, while prepubmed.org also does not have sort functionality, the default ordering appears to be by date. I am unsure how difficult sort functionality would be to implement but some mention of this issue and any plans to implement such a feature in the future is recommended.

3. Outside of benefits to readers, an additional benefit I can see to a cross-platform preprint search engine is that it becomes easier for journal editors to identify/solicit submission of preprints and for grant reviewers to find preprints prior to journal publication. While these are also benefits of preprints in general, which is outside the scope of this article, authors may want to include this additional motivation/rationale as these are also particular benefits of a cross-platform tool. For example, a reviewer may be more likely to read a pending publication if they don't need to search many different sites.
4. It seems the mention of prepubmed.org in the text is limited to a discussion of priority. Since search.bioPreprint is sold as a one-stop shop, I would have instead liked to see a concise comparison with the other cross-platform searches (prepubmed and also Google Scholar) to help users clearly identify any feature differences or benefits of using search.Biopreprint. The unique features of search.Biopreprint are described throughout the article but a concise comparison or table of features/search result data would be advantageous to readers.
5. This article was posted on both F1000Research and bioRxiv. Interestingly, only the bioRxiv version appears as a search result on both search.bioPreprint and Google Scholar, while prepubmed finds both versions. I am curious if some modification by the quality check team would fix this problem or if there is some inherent limitation of search.bioPreprint for preprints posted on more than one server.

Overall, this article clearly describes usage and features of a new tool for cross-platform preprint search that appears to have the advantages of continuous maintenance, useful topic filtering and associated bookmarklet.

Competing Interests: No competing interests were disclosed.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Reader Comment 04 Jul 2016

Jordan Anaya, Omnes Res, USA

Hi, I am the creator of PrePubMed. In response to your comment, yes, PrePubMed sorts articles by date, specifically the date of the earliest version of the preprint that was indexed.

In addition, I would like to comment on search.bioPreprint's ability to perform full text searches. I noticed that the PeerJ Preprints search engine does not perform a full text

search, only title and abstracts. As a result, I think the claim of a full text search should be qualified.

Competing Interests: I made and operate <http://www.prepubmed.org/>, which is another tool for searching for preprints.

Author Response 13 Jul 2016

Ansuman Chattopadhyay, University of Pittsburgh, Pittsburgh, USA

We sincerely thank the reviewer for the constructive review and thoughtful suggestions. Specifically, we appreciate the comments that search.bioPreprint is “a useful tool for the scientific community” and that “the development of a bookmarklet is novel, convenient and very useful.” We submitted a revised version of our paper that addresses the reviewer’s concerns, and we hope the reviewer finds this improved version acceptable without further revision.

Reviewer Comment 1:

There is no mention in the abstract that preprints can be found through Google Scholar. While there is a temporary bug associated with Google Scholar’s treatment of subsequently published preprints (documented in the following blog posts by Dr. Wilke), preprints are indeed searchable across platforms through Google Scholar. Please mention this and note any benefits of your tool. Google Scholar is mentioned on p2 but only in the context of indexing peer reviewed publications.

Author Response:

Thanks for showing us the workaround for retrieving preprint articles via Google Scholar (GS) by adding “preprint” as a text with the search term(s). We missed the preprint search capability of GS as it is not listed as a valid filter such as “include patents.” However, we noticed that the articles retrieved by a preprint search using GS are not always valid preprints. For example, a search for “**asthma preprint**” retrieves many articles already published that were never available as a preprint. When the search results (2,430) are sorted by relevance, the third and fourth citations from the top, “Heliox vs air-oxygen mixtures for the treatment of patients with acute asthma: a systematic Overview” by AMH Ho et al. and “Genomic approaches to understanding Asthma” by LJ Palmer et al., are published articles from Elsevier journal Chest (vol 123, issue 3) and from the journal Genome Research (CSH Press), respectively. After investigating a few non-preprint articles, we found the reason for the inclusion of non-preprint published articles in the GS search result is either a mention of a preprint article in the reference list or in the acknowledgement section.

The results for another search, “CRISPR preprint,” results in **472 articles** by GS and here also the **third citation** from the top is a published paper from the Journal Genome Research (CSH Press).

We added a few sentences on using GS to search for preprints under Conclusions: “**Google Scholar** (GS), a popular scholarly literature search engine that provides cross-discipline search functionality, does not include preprint articles as a filter option. Hence, many avid GS users try a workaround by including preprint with the query term, (E.g., “asthma preprint” or “CRISPR preprint”) with the assumption of retrieving only preprint articles

fetches from major preprint servers. In contrast, the GS search results in a mixed population of articles comprising both actual preprints and peer-reviewed published articles in which the term “preprint” appears somewhere in the full text of the article.”

Reviewer Comment 2:

The search results on search.bioPreprint are ordered by relevance with no option to re-sort. After some time searching using search.bioPreprint, the need to sort results by date became quickly evident. Likely, many users will be interested to know what new preprints are available across platforms since a previous search. A search term on Google Scholar followed by “preprint” and one click on the “sort by date” link produces date-sorted preprints from multiple sources (biorxiv, arrive, peerj etc). Similarly, while prepubmed.org also does not have sort functionality, the default ordering appears to be by date. I am unsure how difficult sort functionality would be to implement but some mention of this issue and any plans to implement such a feature in the future is recommended.

Author Response:

We acknowledge the pressing need for an option to sort preprint articles by date as both reviewers mentioned it. We are actively working to bring this functionality into search.bioPreprint.

We want to stress that the “Sort by date” feature offered by Google Scholar (GS) is abysmal. It drastically drops the number of retrieved articles compared to the default search results. For example, a GS search for “Asthma preprint” retrieves 2,430 citations by the default “Sort by relevance” option, but displays only 6 articles after selecting “Sort by date.” The same thing happens for another search: “Crispr preprint” – Sort by relevance=471; Sort by date=5. In the revised manuscript we mention the need for reordering the search results by date by adding a new paragraph at the end of the Limitations section: “Currently, the search.bioPreprint default search results are ordered by relevance without any option to re-sort by date. The authors are aware of the pressing need for this added feature and if possible will incorporate it into the next version of the search tool.”

Reviewer Comment 3:

Outside of benefits to readers, an additional benefit I can see to a cross-platform preprint search engine is that it becomes easier for journal editors to identify/solicit submission of preprints and for grant reviewers to find preprints prior to journal publication. While these are also benefits of preprints in general, which is outside the scope of this article, authors may want to include this additional motivation/rationale as these are also particular benefits of a cross-platform tool. For example, a reviewer may be more likely to read a pending publication if they don’t need to search many different sites.

Author Response:

Thanks for the suggestion. We added a sentence under Conclusion: “Referees during the grant or journal article review process might also find this bookmarklet useful as it quickly retrieves pre-published articles via the cross-platform preprint search.”

Reviewer Comment 4: It seems the mention of prepubmed.org in the text is limited to a discussion of priority. Since search.bioPreprint is sold as a one-stop shop, I would have instead liked to see a concise comparison with the other cross-platform searches (prepubmed and also Google Scholar) to help users clearly identify any feature differences

or benefits of using search.Biopreprint. The unique features of search.Biopreprint are described throughout the article but a concise comparison or table of features/search result data would be advantageous to readers.

Author Response:

Thanks for the suggestion. Google Scholar does not provide an option to limit searches to preprint articles, and the workaround of including “preprint” with the query term results in a mixed population of articles comprising both actual preprints and peer-reviewed published articles. We are hesitant to consider Google Scholar as a preprint search engine and comparable to search.bioPreprint and PubMed. The purpose of this article is to present search.bioPreprint as a means to locate preprint articles, and its release pre-dates PubMed. We leave it to others to determine the pros and cons of using search.bioPreprint, and hope that they leave comments so we can improve the tool when possible.

Reviewer Comment 5:

This article was posted on both F1000Research and bioRxiv. Interestingly, only the bioRxiv version appears as a search result on both search.bioPreprint and Google Scholar, while PubMed finds both versions. I am curious if some modification by the quality check team would fix this problem or if there is some inherent limitation of search.bioPreprint for preprints posted on more than one server.

Author Response:

We appreciate your efforts in discovering this search error. The Health Sciences Library System’s quality check team has investigated this issue and is working on a solution. We anticipate a quick fix of this problem.

Competing Interests: No competing interests were disclosed.

Author Response 13 Jul 2016

Ansuman Chattopadhyay, University of Pittsburgh, Pittsburgh, USA

Jordan ,

We appreciate your careful reading of our article and clarification of PeerJ’s search limitations.

In the latest version we revised the text to read “We want to acknowledge this new resource, but emphasize that search.bioPreprint offers not only full text searching (with the exception of PeerJ Preprints), but also topical and source-based clustering of results. In addition, our tool has been available since mid-February 2016, around the same time as the [ASAPbio meeting](#), where it was mentioned during discussions.”

Competing Interests: No competing interests were disclosed.

Reader Comment 14 Jul 2016

Jordan Anaya, Omnes Res, USA

Thank you for the change. I also recently noticed that the search engine for arXiv q-bio does not perform a full text search. arXiv recently added an experimental full text search option which is available under "Advanced Search": <https://arxiv.org/find>, however it appears to be limited. Regardless, it seems that search.bioPreprint does not utilize this experimental full text search. Can you comment on your plans for arXiv full text searching?

Thanks, Jordan

Competing Interests: I made and operate <http://www.prepubmed.org/>, which is another tool for searching for preprints.

Reader Comment 20 Jul 2016

Jordan Anaya, Omnes Res, USA

I would like to comment on Google Scholar's ability to search for preprints. As covered by Jessica Polka in this [blog post](#) you can use the advanced search and type in 'bioRxiv OR "PeerJ Preprints" OR f1000Research OR arxiv' into the "published in" field. Google Scholar also indexes PDFs, so it is possible to perform a full text search for journals such as PeerJ Preprints.

Competing Interests: I made and operate <http://www.prepubmed.org/>, which is another tool for searching for preprints.

Reader Comment 24 May 2017

Jordan Anaya, Omnes Res, USA

I would like to clarify and/or raise some issues with this article and accompanying comments.

1. Reviewers Prachee Avasthi and Cynthia Wolberger both emphasized the importance of being able to sort by date, and in response the article was edited to say: "Currently, the search.bioPreprint default search results are ordered by relevance without any option to re-sort by date. The authors are aware of the pressing need for this added feature and if possible will incorporate it into the next version of the search tool."

However, it has been nearly a year and this feature has not been added.

2. The article states: "Until the creation of search.bioPreprint there has been no simple and efficient way to identify biomedical research published in a preprint format..."

This is simply not true as Google Scholar indexes preprints. This was pointed out by

Prachee Avasthi and in response the authors edited the text to include an incorrect method for finding preprints with Google Scholar. In a previous comment I pointed out how to correctly search for preprints with Google Scholar, and it appears the authors read the comment given they utilize the method at this page on their site:

<http://www.hsls.pitt.edu/gspreprints>

3. In his comment the author states: "We want to stress that the 'Sort by date' feature offered by Google Scholar (GS) is abysmal. It drastically drops the number of retrieved articles compared to the default search results."

This feature of Google Scholar is indeed limited, as it restricts the results to articles which were published in the past year. However, if the goal is to find recent preprints then this limitation shouldn't be a problem and I don't know that I would classify the feature as "abysmal".

4. The article states: "As new preprint servers are introduced, search.bioPreprint will incorporate them and continue to provide a simple solution for finding preprint articles."

New preprint servers have been introduced, such as preprints.org and Wellcome Open Research, but search.bioPreprint has not incorporated them.

5. Prachee Avasthi pointed out that the search.biopreprint search engine cannot find this F1000Research article about search.biopreprint. It only finds the bioRxiv version. In response the author stated: "The Health Sciences Library System's quality check team has investigated this issue and is working on a solution. We anticipate a quick fix of this problem."

This problem has not been fixed.

Competing Interests: I made and operate <http://www.prepubmed.org/>, which is another tool for searching for preprints.

The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias
- You can publish traditional articles, null/negative results, case reports, data notes and more
- The peer review process is transparent and collaborative
- Your article is indexed in PubMed after passing peer review
- Dedicated customer support at every stage

For pre-submission enquiries, contact research@f1000.com

F1000Research