



Automated screening of COVID-19 preprints: Can we help authors to improve transparency and reproducibility?

Peter Eckmann, Nico Riedel, Halil Kilicoglu, Gerben ter Riet, Cyril Labbé, Jennifer Byrne, Guillaume Cabanac, Amanda Capes-Davis, Bertrand Favier, Shyam Saladi, et al.

► To cite this version:

Peter Eckmann, Nico Riedel, Halil Kilicoglu, Gerben ter Riet, Cyril Labbé, et al.. Automated screening of COVID-19 preprints: Can we help authors to improve transparency and reproducibility?. Interdisciplinary Meta-research & Open Science Conference (AIMOS 2020), Dec 2020, Sydney, Australia. , 2020. hal-03078870

HAL Id: hal-03078870

<https://hal.science/hal-03078870>

Submitted on 16 Dec 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Automated screening of COVID-19 preprints: Can we help authors to improve transparency & reproducibility?

Peter Eckmann¹, Nico Riedel², Halil Kilicoglu⁴, Cyril Labbé⁵, Gerben ter Riet^{6,7}, Jennifer Byrne⁸, Guillaume Cabanac⁹, Amanda Capes-Davis^{8,10}, Bertrand Favier¹¹, Shyam Saladi¹², Peter Grabitz^{2,3}, Alexandra Bannach-Brown², Robert Schulz^{2,3}, Sarah McCann^{2,3}, Rene Bernard^{2,3}, Anita Bandrowski¹, Tracey Weissgerber^{2,3}
¹UC San Diego, La Jolla, CA, USA, ²Berlin Institute of Health, Germany, ³Charité, Berlin, Germany, ⁴Univ. of Illinois at Urbana-Champaign, IL, USA, ⁵Univ. Grenoble Alpes, France, ⁶Univ. of Amsterdam, Netherlands, ⁷Amsterdam Univ. of Applied Sciences, Netherlands, ⁸Univ. of Sydney, Westmead, New South Wales, Australia, ⁹Univ. de Toulouse, France, ¹⁰CellBank Australia, ¹¹Univ. Grenoble Alpes, La Tronche, France, ¹²California Institute of Technology, CA, USA

Background

- Preprints have grown in popularity since COVID-19 emerged
- Rapid publication is useful during a pandemic, but the lack of peer review has concerned many scientists
- Can we evaluate preprints at scale without relying on authors or the knowledge of readers?

Evaluation

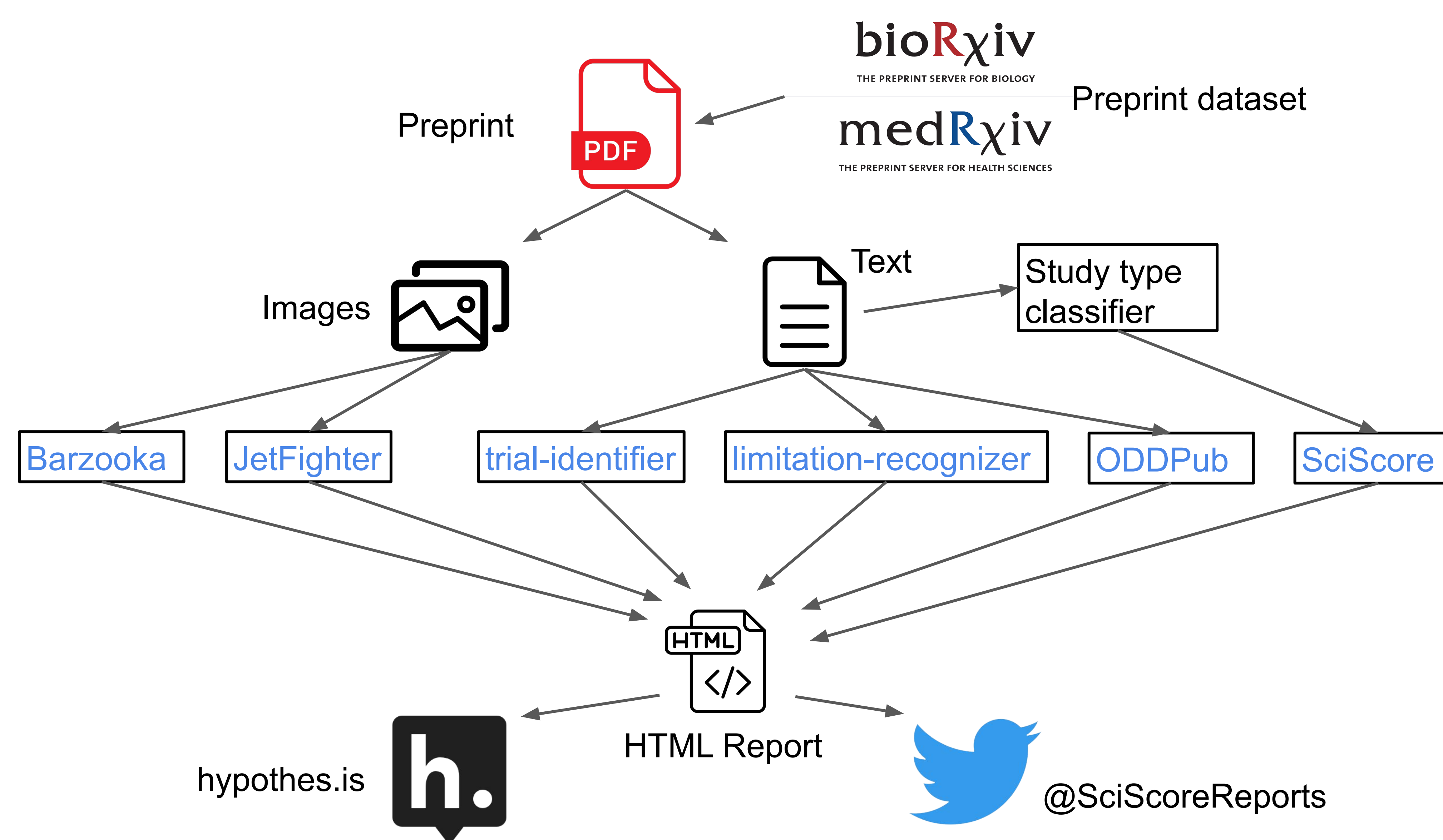
- Goal to automatically evaluate COVID-19 preprints for reproducibility criteria
- Each preprint is downloaded, parsed, and analyzed by a set of tools:
 - SciScore** screens for rigor criteria defined by NIH and resources used (software tools, cell lines, etc.)
 - ODDPub** screens for the presence of open data and code
 - Limitation-recognizer** screens for study limitation statements
 - Barzooka** screens for bar graphs used for continuous data
 - JetFighter** screens for rainbow color maps
 - Seek&Blastn** screens for correctly identified nucleotide sequences
 - Trial-identifier** screens for and verifies clinical trial numbers

Visit us at

<https://scicrunch.org/ASWG>

[@SciscoreReports](https://twitter.com/SciscoreReports)

Pipeline overview



Sample report and replies

medRxiv THE PREPRINT SERVER FOR HEALTH SCIENCES

Phase 1 Assessment of the Safety and Immunogenicity of an mRNA-Lipid Nanoparticle Vaccine Candidate Against SARS-CoV-2 in Human Volunteers

Abstract

There is an urgent need for vaccines to counter the COVID-19 pandemic, due to infections with severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). Evidence from observational data and preclinical studies has identified the viral spike (S) protein as a key antigenic target for protective immune responses. We have applied an mRNA-based technology platform, 'mRNA-Lipid Nanoparticles' (mRNA-LNP), to develop a COVID-19 vaccine candidate encoding for a stabilized form of S protein encapsulated in lipid nanoparticles (LNP). Following demonstration of protective immune responses against SARS-CoV-2 in animal models we performed a dose-escalation phase 1 study in healthy 18-60 year-old volunteers.

Results

This interim analysis shows that two doses of mRNA-LNP ranging from 2 up to 12 µg per dose, administered 28 days apart were safe. No vaccine-related serious adverse events were reported. There were dose-dependent increases in frequency and severity of solicited systemic adverse events, and to a lesser extent of local reactions, but the majority were mild or moderate and transient in duration. Immune responses were measured as IgG antibodies against S protein or its receptor-binding domain (RBD) by ELISA, and SARS-CoV-2 virus neutralizing antibodies measured by micro-neutralization, displayed dose-dependent increases. Median titres measured in these assays two weeks after the second 12 µg dose were comparable to those measured in a comparable cohort from C19-19 vaccine. Serum responses

Conclusion

The vaccine candidate was well tolerated and induced robust immune responses. These findings support the progression to phase 2 studies.

Keywords

COVID-19, SARS-CoV-2, mRNA-LNP, vaccine, immunogenicity, safety

Replies:

Nathalie CHARLOTTE @NathalieCHARLOTTE
Hi SciScore Reports, thank you for checking my article. I submitted the method section including the author declarations to SciScore a few hours ago after reading your tweet. I obtained a 3/5 rigor score (see below).

Ralser Lab @RalserLab
Replying to @SciscoreReports
"0 resources" Hehe this is the most flawed algorithm ever. Its a huge resource paper

Ania Korsunska @akorsunska
power analysis was given in "materials and methods" under the heading "Acquisition, justification, treatment and disposal" section. The text is "N=[z*(2)*p(1-p)]/e^2 = [1.282 * 0.99 (1-0.99)] / 0.052 = 7. N = sample size; Z = the z score, which is 1.28 for power 0.8; p = ..."

Just found out about @SciscoreReports: interesting approach to automating review of scientific articles for rigor and reproducibility. Would this eventually replace human reviewers or just be a useful additional tool? 🤖

scicrunch.org/ASWG

Results

- Study design features
 - 75% of analyzed preprints are secondary analyses, modeling studies, or cell line studies
 - 20% addressed sex as a biological variable, despite known sex differences in COVID-19
 - 6.1% used model organisms, mainly mice
- Transparency
 - 34.4% included self-acknowledged study limitations
 - 14.3% shared open code
 - 13.6% of preprints shared open data
- Data presentation
 - 7.6% used rainbow colormaps, which are not colorblind safe and can create visual artifacts for viewers with normal color vision
 - 7.3% used bar graphs for continuous data, which can lead to misleading figures
- Combined, the automated Tweets have been viewed about 380,000 times
- Current average of ~1,000 views and ~10 link clicks per day
- The account has accumulated a total of
 - 2459 link clicks
 - 98 retweets
 - 42 replies

Conclusions

- It is feasible to conduct large-scale automated screening of preprints for common quality criteria and provide feedback to study authors and readers before publication
- Reports can publicly raise awareness of factors that affect study quality and reproducibility, while helping authors to present their research in a more transparent and reproducible manner.