

## 第 2 回 SPARC Japan セミナー2020

「プレプリントは学術情報流通の多様性をどこまで実現できるのか？」

# 生命科学における プレプリントや SNS 活用の現状と課題

坊農 秀雅

(広島大学 大学院統合生命科学研究科 生命医科学プログラム)

### 講演要旨



bioRxiv (バイオアーカイブ) をはじめとするプレプリントの活用が生命科学分野においても急速に広まり、演者も海外での学会発表の前にはその発表のプレプリントをアップロードするようにしている。また、ResearchGate に代表される研究者向けの SNS も普及し、それを介した (電子的な) 別刷り請求やポストドク申し込みなど、従来の学術コミュニケーションに変革が訪れている。しかしながら、それぞれその情報の質や有償コンテンツの受け渡しなど問題点も多く指摘されている。そこで本講演では演者の実体験を交えながら、その利点と欠点に関して生命科学分野におけるプレプリントと SNS の使われ方の現状を紹介する。

### 坊農 秀雅

理化学研究所、埼玉医科大学を経て、2007年より情報・システム研究機構ライフサイエンス統合データベースセンター (DBCLS) にて統合TVや統合データベース講習会AJACSを立ち上げ、日本DNAデータベース (DDBJ) と連携して第きび塩基配列データ解析技術開発とその利活用を行う。2020年4月より広島大学大学院統合生命科学研究科 特任教授として卓越大学院 (ゲノム編集先端人材育成プログラム) でのバイオインフォマティクス教育に取り組みつつ、ゲノム編集データ解析基盤技術の開発とバイオインフォマティクスによる遺伝子機能解析を行う研究室 (bonohulab) を立ち上げている。京都大学博士 (理学)。



私は以前、この SPARC セミナーの運営委員をしており、ちょうど 2 年前 (2018 年 10 月) にプレプリントの話をしたので、この 2 年でどう変わったかというところに注目して聞いていただければと思います。今日はプレプリントと研究者版 SNS のお話を、私の事例を使ってご紹介します。

私は 2020 年 4 月に広島大学に移りました。ゲノム編集 (Genome editing) を使うセンターにも所属しています。ゲノム編集の実験を行うには、その前段階でさまざまなデータを集めなければいけません。例えば、どの遺伝子を狙ってゲノム編集をするかを決めたり、ヒトではなく新しい生物の場合は、ゲノム配列を読ん

で遺伝子を決めたりします。そのためのゲノム編集データ解析基盤技術の開発と、バイオインフォマティクスの遺伝子機能解析研究を、新しい研究所を立ち上げながら進めています。

### bioRxiv

図 1 は 2 年前に出したスライドです。私は公共のデータベースの再解析 (メタ解析) を行うための手法を論文を作って出したのですが、2 年前にその話をしました。当時、私は静岡に住んでいて、2018 年 2 月 23 日の富士山の日とその論文のプレプリントを bioRxiv に出して、8 カ月後の 10 月にプレゼンをしました。

その 8 カ月の間にアブストラクトの閲覧が 907 回、PDF のダウンロードが約 300 回あり、たくさん見られるようになってうれしいという話をしました (図 2)。その後どうなったかという、2 年後の 2020 年 11 月末現在でアブストラクトの閲覧数は 3,000 を超え、PDF のダウンロード数も 1,000 を超えています (図 3)。また、プレプリントでアップするときは PDF を論文の体裁にしてアップするのですが、bioRxiv のシステムが良くなり、そこからテキストマイニングをして、普段ジャーナルで見るとようなフルテキストへの組み替えも自動的にできる仕組みに変わりました。これは medRxiv もそうです。

その他の違いとしては、「これはプレプリントなので注意して見てください」という赤枠の注意書きが付きまして。それから、左の真ん中あたりのだいたい色で囲んでいるところですが、査読論文が通ると自動的にリンクが張られるようになりました。プレプリント

で放ったらかしていても、古い DOI のものとトレースできる仕組みになっています。ですので、これはもう古いのではないかという懸念は少ないのではないかと思います。ただ、もちろん漏れがあるので、プレプリントのままで、査読の方に出ているのに見つけれないということはあるかもしれません。

そういう状況で、bioRxiv のシステムも社会も 2 年で変わりましたが、研究者社会の方はあまり変わっていません。プレプリントを出すメリット・デメリットは全く変わってなくて、残念ながら、研究者でも一部の人間だけが先取制を担保するために出すということが多くです。多くの保守的な日本の研究者は、「プレプリントはよく分からないからいいや」という状況です。

変化したのは、medRxiv が COVID-19 の影響で広く知られるようになったことです。これはもちろんいいのですが、査読論文との違いを無視したり、プレプリントに関する無理解が露呈したりする状況があります。また、一部の共同研究者は、査読前に bioRxiv に出し、より早く「自分たちがこの研究をしたのだ」ということを言い、エディターの人たちが悪いことをして別の論文を先に通すということがないようにしています。早く bioRxiv に出せば、自分たちが先にやったということが分かるので、そのようにする人が増えてきています。

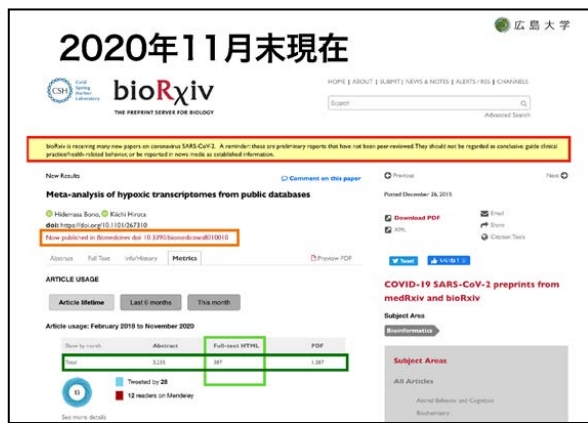
しかし、プレプリントが増えている中で、bioRxiv で受け入れてくれないタイプの論文もあります。例え



(図 1)



(図 2)



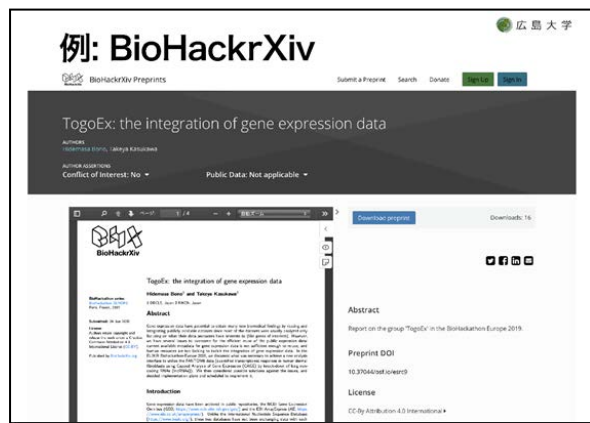
(図 3)

ば、私は生命科学の中でも特にバイオインフォマティクスというプログラムを書いたりするタイプの研究をしているのですが、みんなで集まって Hackathon (hack と marathon をかけた言葉) というミーティングのようなものをしたときのレポートを私たちは査読論文誌にずっと出してきたのですが、そのプレプリントを bioRxiv に出したら「ミーティングレポートは受け付けません」と断られました。

それ以外の変化として、論文誌を出しているシュプリンガー・ネイチャーや MDPI において、論文投稿と同時に自分たちのところがやっているプレプリントに投稿されるような仕組みができ、プレプリントに誘導する仕組みがだいぶ出てきています。有名なのは、今月に入ってから出た eLIFE というものです。割と良い論文誌を出すということで研究者の間では「ここに出したい」という人が増えています。その eLIFE が、"publish, then review"ということで、まず論文をパブリッシュして、その後レビューを行うということを始めるとアナウンスしています。これも結構話題になっており、このような F1000Research モデルが広まっていくといいなと思って見ているところです。

### BioHackrXiv

先ほどの Hackathon のレポートは bioRxiv で受け入れられなかったのが、BioHackrXiv に出しました (図 4)。これもプレプリントで、DOI は付くし、ORCID などにもレコードが入りますが、Europe PMC には載



(図 4)

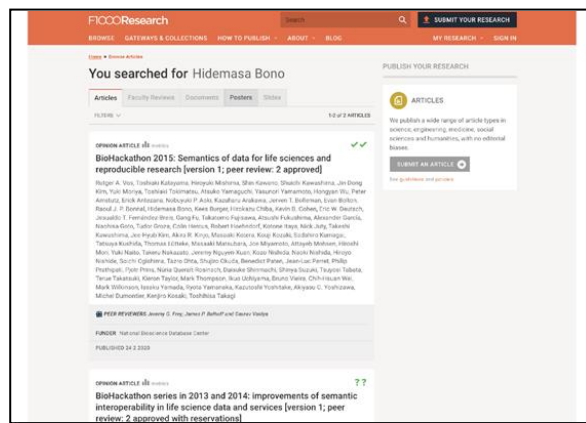
っていません。こういう形で、自分たちが BioHackathon のミーティングでまとめたレポートをプレプリントとして論文に出すことができます。これは 2019 年の秋に私どもがヨーロッパの BioHackathon で議論したことをレポートにして出したものですが、こういう形で残していくことができるのは、それはそれでいいなという例の一つです。

### F1000Research

図 5 は F1000Research で、これも私が出したものです。実は BioHackathon のレポートとして 2 回、F1000Research に出して、上と下で別々の論文ですが、上の 2015 はピアレビューで OK となっていて、2 人分のチェックが付いており、PubMed で検索できるようになっています。下の 2013 and 2014 は「??」と付いています。これは、レビューされたがアプローブされず、PubMed などにも載らず、このままの状態でのプレプリントのように F1000Research のサイトにあるという状態です。

### ResearchGate (RG) : 研究者向けの SNS

ResearchGate (RG) は研究者向けの SNS です。割と古くて、2008 年にボストンで設立され、2013 年からはドイツのベルリンに本部を移しています。少し古いデータですが、2016 年の公称利用者数は 900 万人です。いつから自分が登録しているかという詳しいレコ



(図 5)

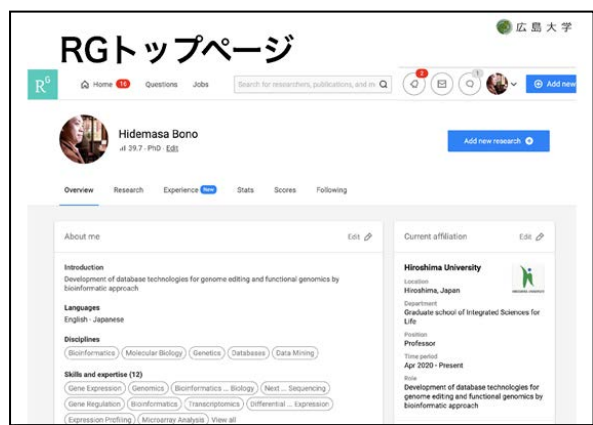
ードは残っていませんでしたが、私はどうも 2015 年ごろから登録して使っているということが分かりました。

基本的には researchmap や ORCID のように、研究者の論文業績リストのような感じかと思っていましたが、ORCID と同様、この ResearchGate もどんどん自動で登録されていくようです。誰かが登録しないといけないようですが、共著者のうちの誰かが登録すれば他の人にも行くという仕組みが古くからできていたようで、自動で入るような印象があります。researchmap も最近バージョンが上がってそういう仕組みになってきているので、大して変わらないとは思いますが、そういうことを早く取り入れたというイメージです。

それから、following や followers という概念もありますが、Twitter のようにフォローしていないと見られないということではなく、どちらかという研究に関係があるキーワードや誰かの名前で検索して見に行くような感じです。全然知らない、フォローしていない人もやりとりができ、実際のところ followers や following というのはあまり関係ない仕組みになっています。

私がいいと思っているのは、業績に対する統計値の可視化ができることです。自分たちの論文が引用された・されていないということが 1 週間ごとの解像度で出るので、割とリアルタイムにそういうものが見えて面白いです。

図 6 は私自身のトップページです。名前や PhD な



(図 6)

どが書いてあります。右上にあるアラートは、Facebook のアラートと同じですが、鐘マークのところに 2 と書いてあります。クリックすると、自分の論文を誰かが新しく引用したというお知らせが来ています。放っておくとすぐたまってしまうので、自分のトップページを見ろよという誘導がうまくできていると思います。

右上の会話マークはチャットのようなものです。グレーで 1 と書いてあるのは、見たけれどもずっと残してあるということです。具体的には「論文の PDF を下さい」というリクエストが来ています。そういうものも全部ここに出ます。

SNS の機能もちろんあります。私自身は 4 年間で 10 個以下のメッセージしかありませんが、ほとんど国外からのメッセージです。一番多いのが別刷りの PDF の請求です。私はオープンアクセスの論文にばかり出しているの、よく論文を読んでいないのだからと思いつつ URL を教えてあげて解決しているのですが、中には「オープンアクセスではない論文誌の PDF が欲しい」というメッセージも来て、これはどうなのかと思っています。そこは個人によって判断が分かれるところで、私は、自分がファーストオーサーやコレスポンディングオーサーでないものは断っています。あとは雑誌特集号への寄稿依頼や、最近増えてきたのはポスドク申し込みです。そういうものがメッセージとして送られてくることが多いです。

また、researchmap 的な機能として、研究業績が出ます。特徴的なのは、その論文のページを何回見たかということが集計されることです。論文を登録する際に、普通の査読論文だと DOI を入れれば勝手に登録されますが、プレプリントの場合は論本文体の PDF をアップしないと登録されません。CC-BY などになっている普通の査読論文だと、勝手に収集されてここに張られていくようです。

そして、スコアが研究者ごとに付いています。このスコアは、論文を出した数、誰かからの質問に答えた数、followers の数などで出るので、私の場合はほ

とんどが論文で数えられています。

それから、h-index も同じように計算されています。これに関しては、ほとんど Google Scholar と一致しており、ある程度信頼できるのではないかと思います。

図 7 は 1 週間ごとの統計値です。これは 11 月と 10 月ですが、誰かが見に来た回数、フルテキストを読んだ回数、Citations、あとは Recommendations ということで、誰かがその論文を Recommend するとその数も出るので、私の研究だとそれはあまり出ていません。Research Interest など、そういうものを計算して ResearchGate 的にスコア化しているものだと思います。これらは weekly 以外に monthly でも見られます。そういう可視化の機能がいいと思っています。

その他の機能としては、自分が出した論文を研究のプロジェクトとしてまとめることができるので、頭の整理になっていいということもあると思います。

それから、海外のものばかりですが、宣伝情報や job 情報がメールで来たり、本の広告が JREC-IN 的な感じで来たりします。

## まとめ

以上、プレプリントと ResearchGate (研究者版 SNS) のお話をしました。両方とも日本の研究者がほぼ全員使っているということではなく、まだ一部の人が便利だと言って使っている感じです。しかし、フルテキストは ResearchGate に、プレプリントは Google にインデックスされていますから、ググってとり着き、知

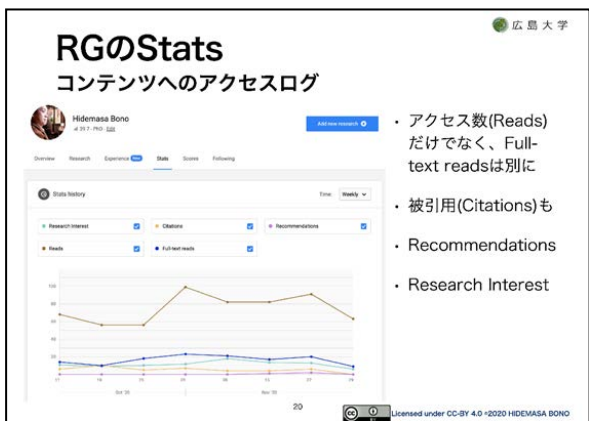
らないうちに使っていたという例は結構あるようです。

しかし、研究者は Twitter を匿名 (のつもり) で使っている人が多いです。実名にしているのは、今後の評価などに影響すると考えているからのようです。そういう状況なので、やはりプレプリントは研究費を出している側で「出してください」と義務付けるべきだと思います。研究費を出す海外の機関、例えばウェルカム・トラストは、独自にプレプリントのサーバーを用意し、「そこに入れてください」というようにしていますし、National Institutes of Health (NIH) の PubMed Central (PMC) も、NIH グラントでやったものは自分たちのところに入れることになっていたと思うので、日本もトップダウンでそのようにしていった方がいいと個人的には思っています。

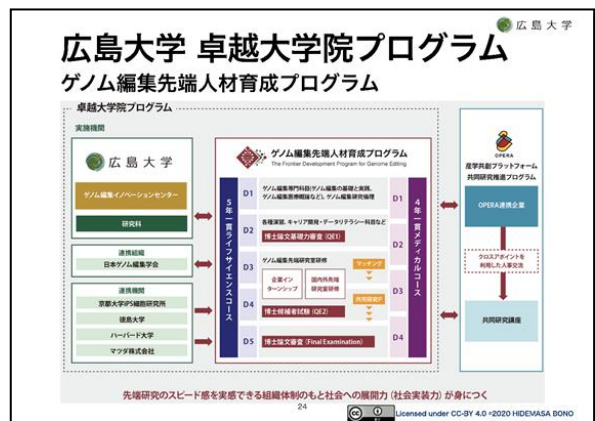
一番問題なのは、査読論文しか評価しないという現在の基準だと思います。研究者は割と皆さんそういう頭があり、「査読論文しか出さない。他はやっても労力の無駄だ」と思っていることが多いので、そういう意識を変えていくような活動を若い世代に対して行っていく必要があると思います。そして、やはりトップダウンの強制力が必要です。これも前から言っているのですが、なかなか日本では実現しません。

## 広島大学 卓越大学院プログラム

そのようなことのための教育を、現在、広島大学の卓越大学院プログラムでさせていただいています (図 8)。ゲノム編集先端人材育成プログラムの中でバイオ



(図 7)



(図 8)

インフォマティクスの講義を担当し、コンピューターを使ったデータ解析や、データサイエンス関係で生物の研究を進められる人材の育成を行っています。

●矢吹 先生方とお話をしていると、ResearchGate を使っている方がいらっしゃる一方で、全く使わない方もいらっしゃいます。使っていると、もうこれだけでいいという感じになってくる先生もいらっしゃって、人によっては相当使い勝手がいいのだらうなという印象を受けていますが、プレプリントの話を含めて、ResearchGate が研究活動自体にうまく活用できている部分は何かありますか。

●坊農 やはり自分から情報発信しないと集まってこないという部分はありますが、私が自分の研究テーマで論文を出していると、特集号などのお知らせが来て「それに出してくださいね」という依頼が来ることがあるので、それはプラスになっているかなと思います。興味がある人が質問してくるということがあればいいのですが、私に関しては今のところそれほどなくて、「ポストドクに来たいです」という人が来るくらいです。しかし、人によってはここでそういう研究交流が生まれて、そこから共同研究が始まることあるのではないかと思います。私は ResearchGate ではなくて、Twitter の日本語でのやりとりで共同研究が始まったことは割とありまして、むしろ Twitter 世代です。まだ利用者が多くないときの Twitter でそういうやりとりがあり、プラスになっています。

●矢吹 ありがとうございます。ちょうど私も大学の研究力分析を行っており、URA という立場もあって、研究活動と SNS に関心があったので伺いました。

コメントが一つ来ています。「政府によるプレプリントの義務化あるいは著者最終稿の公開義務化には大賛成です。ぜひパネル討論で議論してください」とい

うことなので、この後、パネルで言及していただければと思います。

もう一つ、「プレプリント論文におけるデータ公開の実例をご紹介いただけますか」という質問が来ています。

●坊農 プレプリントの場合は、プレプリントを出したときにデータも全部出さないといけません。査読論文の場合は査読論文の出版時に出していたのが、先になるということです。私自身は塩基配列、DNA の配列などを決めてそれを出すのですが、国際塩基配列データベースに登録し、その登録番号を論文に含めないといけません。査読中はそういうものは伏せておく、あるいは reviewer だけが見られるリンクを与えるのですが、arXiv などのプレプリントの場合は、アップロードした瞬間にそれが公開されていないといけません。それで、論文が出たときに、既にデータも利用可能になっているという状況が大事なのです。

今言ったのは塩基配列のデータベースの話ですが、定型の塩基配列のデータだけではなく、論文に付随したさまざまな種類のデータを、例えば Figshare や datadryad などにアップして、その DOI を論文に書いておき、そういうものは無制限に誰でもアクセスできるようにしておくことがプレプリントでは必須とされています。しかし、なかなかそれも広まっておらず、データはまだ公開していないけれども取りあえず自分たちの先取性のためだけに bioRxiv などのプレプリントにアップしているという例がたまにあるようです。ぜひ皆さんにプレプリントのルールを知っていただき、データがより公開・再利用されるような研究社会になればよいと考えています。