May 30th, 2024

Dear editorial committee of Nature Microbiology,

We are pleased to submit our manuscript entitled “**Pre-trained molecular representations enable antimicrobial discovery**” for consideration for publication as an **Article** in Nature Microbiology. In the present study, we have developed, and experimentally validated, a machine learning framework that enables predictions of antimicrobial activity for any chemical structure, against various bacterial species*.* Our addresses the need for predictive models in domains where there is a lack of large-scale data resources and can be a valuable tool to the research community.

The prevalence of resistance to current antibiotic treatments is a major health concern that is set to increase in the coming years. Deep-learning-assisted strategies can help prioritize compounds for experimental validation, thereby increasing the rate at which novel antibiotics are found. However, current strategies require large amounts of training data and a publicly available large-scale data resource for antimicrobial discovery is not yet available. To address these challenges, we developed a two-stage deep-learning strategy that enables researchers to assess the antimicrobial potential of any compound of interest against various bacterial species:

1. To address the lack of large training datasets, we introduce a self-supervised pre-training framework (called MolE) that learns a task-independent representation of chemical structures. This representation captures chemically relevant features and enables machine-learning models to make accurate property predictions.
2. We combine the pre-trained MolE representations with measurements of growth-inhibitory activity from [Maier et al., 2018](https://www.nature.com/articles/nature25979), to create model that scores compounds with respect to their antimicrobial potential.

Using our framework, we prioritized five human-targeted drugs for experimental validation and confirmed three of them as novel growth-inhibitors of *Staphylococcus aureus*. We are also able to build on the results from studies such as [Stokes et al., 2020](https://pubmed.ncbi.nlm.nih.gov/32084340/), by re-discovering structurally novel antibiotics using our framework. In this way, our study offers a strategy that incorporates recent advances in the field of deep learning to tackle chemical and biological data scarcity, thereby accelerating antimicrobial discovery.

We believe that our framework will be a valuable tool for a wide range of microbiologists who wish to quickly assess the spectrum of antimicrobial activity for their compounds of interest. Furthermore, the MolE representation can be used directly in combination with widely accessible machine-learning methods for various property prediction tasks. For these reasons, we consider that our work will be of great interest to the readership of Nature Microbiology.

Should our manuscript be considered for peer review, we would like to suggest the following reviewers: Jonathan Stokes (McMaster University), Zhonglin Cao (Carnegie Mellon University), Patrick Aloy (Institute for Research in Biomedicine Barcelona) and Elhanan Borenstein (Tel Aviv University).

All authors have read and contributed to the content of this manuscript. A pre-print of our work is available at BioRxiv (https://www.biorxiv.org/content/10.1101/2024.03.11.584456v1). This work is original, has not been published elsewhere, and is not being considered for publication at other venues.

We look forward to hearing from you regarding your decision to consider our manuscript.

Sincerely,