KOCH INSTITUTE
for Integrative Cancer Research at MIT

Massachusetts Institute of Technology ● ki.mit.edu
77 Massachusetts Avenue ● 76-361 ● Cambridge, MA 02139

Francisco J. Sanchéz-Rivera, PhD
Assistant Professor, Biology
Massachusetts Institute of Technology
617-715-3389
fsr@mit.edu

July 17th, 2024

Barbara Cheifet, Ph.D.
Chief Editor, *Nature Biotechnology*
Nature Publishing Group

Dear Dr. Cheifet,

Following our conversation on June 12th, I am writing to submit our manuscript *"Computational modeling of human genetic variants in mice"* for consideration as a *Brief Communication* at *Nature Biotechnology*.

Many human diseases are linked to genetic alterations whose roles in pathogenesis remain unknown. For instance, tumors harbor a **complex spectrum of point mutations and chromosomal rearrangements** that can contextually perturb gene function in different ways. Understanding how these mutations impact diseases like cancer is important to advance our knowledge of human health and develop more precise therapies.

Cancer genetics research often relies on genetically engineered mouse models (GEMMs) expressing well-defined oncogenes (e.g. *Kras*^{G12D}) or lacking key tumor suppressor genes (e.g. *Trp53*). While powerful, most GEMMs do not recapitulate the precise genetic lesions found in human cancer and other diseases, thereby limiting their predictive potential. New **precision genome editing technologies** like **base editing** and **prime editing** have started to overcome these issues, yet several important challenges remain unsolved.

Despite extensive genetic homology between humans and mice, accurate modeling of human genetic variants in the mouse genome is a complex multifactorial problem. **Species-specific genetic inconsistencies present a significant challenge** that complicates the development and benchmarking of GEMMs for accurate studies of human genetic variation. Addressing this challenge is particularly urgent in the current genome editing era where researchers can engineer and interrogate virtually any type of genetic variant in diverse experimental models, including GEMMs. Engineering the right variant is arguably the most important step.

We tackled these challenges by developing **H2M (human-to-mouse)**, a computational pipeline to analyze human genetic variation data to accurately model and predict the functional consequences of equivalent mouse variants. **H2M integrates mouse-to-human and paralog-to-paralog variant mapping analyses with precision genome editing pipelines to devise strategies tailored to model specific variants in mice**. We used H2M to construct a database containing > 3 million human-mouse equivalent mutation pairs, as well as ready-to-use *in silico*-designed base and prime editing libraries to engineer 4,944 recurrent variant pairs. These studies represent **the first dedicated effort to bridge human genetics, mouse models, and scalable precision genome editing** by developing and applying computational methods for cross-species functional genetic analysis and genome engineering.

H2M is not just a computational tool for cross-species variant mapping and design of precision genome editing reagents. Instead, we show that **H2M is a multidirectional generator of genetic variant information**

that is designed to also perform reverse mouse-to-human mapping and other types of functional interspecies modeling using other model organisms.

The rich datasets generated by H2M are thus quite versatile, and we believe they can be leveraged for **computational identification and experimental modeling of neoantigens conserved across ≥ 2 species**. For instance, we found that predicted variant pathogenicity and immunogenicity scores are highly correlated between human-mouse variant pairs, suggesting that variants with similar sequence change effects may also exhibit broad interspecies functional conservation. This is a particularly exciting hypothesis that could be systematically tested by integrating high-throughput methods like EpiScan and TCR-MAP to interrogate **thousands of candidate mutations predicted by H2M to generate immunogenic peptides**. New GEMMs expressing tagged endogenous versions of MHC-I alleles, including *H2-K1* (*K^{b}Strep* knock-in mice), could also be used to engineer and study high-priority antigens predicted to exhibit cross-species functional conservation and immunogenicity. These results establish the potential of integrating cross-species computational analysis with GEMMs to discover, predict, and evaluate the immunogenicity of disease-associated mutations to accelerate neoantigen discovery and precision medicine.

Overall, H2M fills an important gap in the field by establishing a **robust and versatile computational framework to identify and model homologous variants across species** while providing key experimental resources to augment functional genetics and precision medicine applications. The H2M database (including software package and documentation) can be accessed at https://human2mouse.com.

For reviewers, we suggest Luke Gilbert (UCSF, Arc Institute, luke.gilbert@ucsf.edu), leader in high-throughput human and mouse CRISPR functional genomics, Maria Jasin (MSK, jasinm@mskcc.org), internationally-recognized genome editing pioneer, Chris Vakoc (CSHL, vakoc@cshl.edu), leader in CRISPR-based methods for protein domain characterization and paralog discovery, and Kris Wood (Duke, kris.wood@duke.edu), leader in functional genomics and paralog druggability.

The entirety of the material presented in this manuscript has not been published or is under consideration for publication elsewhere, in any form. The newest version of H2M (1.0.3) is available on PyPI at https://pypi.org/project/bioh2m/ and on GitHub at https://github.com/kexindon/h2m-public. A Tutorial for using H2M is also provided in the GitHub repository, in addition to all analysis scripts and codes for figure generation. Further documentation and installation instructions for PEGG are available at https://h2m-public.readthedocs.io.

Please let me know if you have any questions. We look forward to hearing from you soon.

Sincerely,

Francisco J. Sánchez-Rivera