STATISTICAL METHODS

Giving mouse models a better FIT

Normand, R. et al. Nat Methods 15, 1067-1073 (2018)

A typical preclinical experiment will start with mice. A model of a desired disease will be developed, a treatment applied, and any changes in gene expression recorded for comparison against healthy control animals. The gene with the greatest change in expression often becomes the target for further research on the path to human applications.

If only it were that easy. Translational failures between mouse and man abound. But so to does data. Rather than trying to build a better mouse, Shai Shen-Orr and his lab are working to bridge the species gap with computers. "We're living in the 21st century and there's at least 20 years of gene expression and –omic data out there," says Shen-Orr, a computational biologist at Technion-Israel Institute of Technology. From his perspective, a systematic, datadriven way to make sense of all that information was missing.

So recently, he and his lab, along with collaborators at Stanford University,

took a machine-learning approach and developed Found in Translation (FIT). FIT is a statistical model designed to take prior knowledge about the differences between mouse and human biology into account when interpreting experimental gene expression data. Using publically available data in the NCBI GEO repository, the team built a compendium of mouse-to-human gene expression in which they paired mouse model data with human disease data for 28 different diseases. For novel mouse experiments, the FIT model calculates a per-gene effect size based on the relationships in the compendium and predicts a new, absolute effect size researchers could expect to see in a human with the same disease.

The model can 'rescue' mouse genes that might not otherwise have made the cut for further consideration, says Shen-Orr; conversely, it can de-prioritize genes that are differentially expressed in the mouse but that might not necessarily be the most relevant in the human condition.

Though the current iteration works better for some diseases than others, FIT could identify ~20–50% more potentially human-relevant genes than looking solely at raw mouse data. The authors give the example of *ILF3*, a gene that the model predicted should be upregulated in the colons of patients with intestinal bowel disease (IBD) but that hadn't been noted before in mouse or human studies. When they tested new human colon samples, they observed an increase in *ILF3* in patients with IBD, compared to healthy adults.

There's an R package available for those with an interest in the code, as well as a web service that's free for anyone to try at http://www.mouse2man.org.

Ellen P. Neff

Published online: 7 January 2019 https://doi.org/10.1038/s41684-018-0228-z

Poking hurts. **Be nice**.



Cheek and retro-orbital bleeds, tail sticks, sublingual sampling – ouch! Instead, take precise, pure blood samples with low stress and no needles. Nice.

Learn how at www.instechlabs.com/be-nice

