

MACHINE LEARNING

An algorithm to assess what's essential for the developing mouse

Tian, D. et al. *Dis. Model Mech.* **11**, dmm034546 (2018)

Some genes are so essential for development that an organism simply can't live without them. Researchers have now developed a machine learning algorithm that can make that call about any of the 20,000 protein-coding genes in the mouse genome with approximately 80 percent accuracy.

The tool can provide guidance to researchers who are considering knocking out a particular mouse gene to study its function, says Kathryn Hentges, a geneticist and developmental biologist at the University of Manchester who led the work. If the gene is predicted to be essential, "then maybe you want to make a conditional knockout rather than an absolute knockout," she says, so that the animal is viable for study. And because there is a strong overlap between essential genes in mouse and in human, the tool can also help interpret human sequencing studies, Hentges adds. Non-essential genes, for example, are less likely to be involved in developmental disorders.

To develop the tool, Hentges and her colleagues tested several machine learning algorithms on an initial set of mouse genes whose essentiality was already known. An algorithm called "random forest" performed best with their training set. Random forest essentially generates a series of decision trees based on about 100 features it identifies in the genes. Those features don't necessarily have functional value; they are simple yes or no questions, such as "Is the gene longer than 2,000 amino acids," or "Is the gene's protein product localized subcellularly," or "Is the percentage of a particular amino acid in a protein below a certain amount?"

The researchers then applied the algorithm to novel sets of genes. Meanwhile, results from the International Mouse Phenotyping Consortium—which is creating and characterizing knockouts of every protein-coding gene in the mouse—were emerging, providing experimental validation of the tool's predictions. "It was really

valuable to have the IMPC data coming along experimentally at the same time we were doing this," Hentges says.

Although the algorithm's accuracy is far from perfect, "it provides more information than trying to integrate all the data from different sources yourself," Hentges says. Also, she adds, the algorithm is unbiased, so it won't give undue weight to features based on faulty assumptions.

Her team is now developing a similar tool to identify mouse genes involved in kidney disease. That task will likely prove more difficult, Hentges says. Essentiality has a clear phenotype—embryos die—that guides how the algorithm segregates genes. But the phenotypes relating to kidney disease are less clear-cut, which will make it tougher for the computer to seek patterns in the data.

Alla Katsnelson

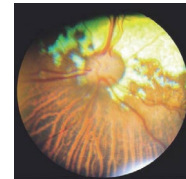
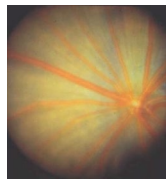
Published online: 18 February 2019
<https://doi.org/10.1038/s41684-019-0251-8>



WHO PERFORMS EYE EXAMS FOR YOUR STUDIES?



Board certified veterinary ophthalmologists are uniquely qualified to consult in the development of the experimental design, including the species selected, appropriate diagnostic tests, and frequency of exams. Coordination between the testing agency and the board certified veterinary ophthalmologist is essential throughout the process, to include protocol development, Standard Operating Procedures (SOP), and assessment of the outcome of testing.



Contact information for ACVO consultants: office19@ACVO.org

ACVO.org

PO Box 1311

Meridian, ID 83680

208-466-7624