

## GENOMICS

**The Black 6 gets company: *De novo* genomes assembled for 16 more mice**Lilue, J. et al. *Nat. Genet.* **50**, 1574–1583 (2018).

Not long after the initial publication of the human genome sequence, the Mouse Genome Sequencing Consortium followed suite with the first draft of a similar resource for the lab mouse (*Nature* **420**, 520–562; 2002). The mouse reference genome has been referenced thousands of times and improved and expanded over the ensuing years, but there's still more work to be done in the realm of mouse genetics.

Like including more mice. “That was a revolutionary resource for the community,” says Thomas Keane, a bioinformaticist at the European Bioinformatics Institute. “But that's just one single strain.”

The reference mouse the murine research community has looked to all these years is the C57BL/6J, a popular strain used in research labs around the world. But the Jackson Laboratory's ‘Black 6’ is hardly the only mouse out there. As of November 6, 2018, Mouse Genome Informatics (MGI), an online resource for researchers looking for information about laboratory mice, lists 49,420 different mouse strains.

Even though lab mice are all the same species, *Mus musculus*, different strains don't have exactly the same genetics. They've been developed by researchers to produce distinct phenotypes, the observable traits that are determined by interactions between an organism's genes and its environment. To fully understand the link between genes and a phenotype, it's important to consider what's unique in the genome of a given mouse. To help, Keane and his colleagues had been creating variation catalogues for different mouse strains against the reference Black 6 for several years. “We were taking the sequencing reads from different strains, placing them onto the reference genome, and just looking for differences,” he explains. “Say, single nucleotide changes, where an ‘A’ changes to a ‘T’.” But that approach can miss novel details, particularly in strains that deviate significantly from the reference.

So Keane, along with an international team of collaborators with diverse expertise in genetics and genome sequencing, decided to start from scratch. They took



Credit: Peter Ginter / Science Faction / Getty

sixteen commonly used lab mouse strains—twelve inbred and four wild-derived—and produced *de novo* reference genomes for each. Those draft genomes are now available online for interested researchers in Ensembl, the University of Southern California's Mouse Genome Browser, and the MGI; the details about the effort are reported in a paper in *Nature Genetics*.

“The main benefit of this paper is that we now have much greater detail on the extent of specific genetic variation for many inbred mouse strains commonly used for laboratory research,” says Kent Lloyd, a veterinary scientist and director of the Mouse Biology Program at the University of California Davis. “This is a significant improvement over previous catalog of SNPs and other genetic variations compared to the C57BL/6J reference genome. The more detailed and extensive genetic diversity provided by this new study greatly informs and begins to explain the diversity of strain-specific phenotype variation.”

Across the sequenced strains, there can be quite a bit of diversity relative to C57BL/6J that hadn't been observed before. “You start to see these completely different gene structures that we just didn't know about... You see new exons, you see new re-arrangements of genes,” Keane says. “The big surprise is just the number of those.”

That underlying genetic diversity has implications for even simple aspects of

studies, such as designing PCR primers, CRISPR targets, or basic assays. Without a genome specific to the strain in question, there can be a bit of guesswork involved, he says. “We've shown examples where if you're using the Black 6 sequence you can potentially get incorrect results.”

There are functional consequences to consider too. Much of the variation observed was found in regions of the mouse genome that contribute to immunity, pathogen defense, and sensory function. “These are genes that are potentially quite important if you're using mouse models to study human disease,” says Keane.

The new sequences revealed a completely novel gene as well, and a large one at that—nearly 6000 amino acids long. “It's present in all the strains, it's present in the Black 6, but it just hadn't been discovered in the earlier rounds of annotation of the reference genome,” explains Keane. Using CRISPR, the team knocked out the gene *in vivo* and observed deleterious effects on brain development in the mutant mice.

Are there still surprises waiting to be discovered in the lab mouse? Possibly. “The draft genomes that we produced are better than what we had before, which was no genomes,” says Keane. “We absolutely know that there are many of these divergent regions that require further sequencing to fully resolve.” But filling in further details is all part of the plan—Keane has funding to upgrade these draft genomes with “third generation” sequencing platforms in 2019.

“This is a first draft, so there's more to come,” says Lloyd. “Nonetheless, researchers will now have access to genetic evidence in context to select the most appropriate inbred mouse strain for their specific research purposes, rather than grabbing the C57BL/6J mouse off-the-shelf out of convenience.”

Ellen P. Neff

Published online: 19 November 2018  
<https://doi.org/10.1038/s41684-018-0197-2>