

Mouse phenotyping sheds light on rare disease

When the first mouse knockout studies were published in 1989, it was hard to imagine that mouse models would be used systematically to functionally annotate the human genome. A recent study in *Nature Genetics*¹ shows that the International Mouse Phenotyping Consortium (IMPC) is well on its way to doing just that. In collaboration with the European Bioinformatics Institute, the IMPC is measuring the broad phenotypic effects of knocking out mouse genes, one at a time, to uncover their biological function. So far, they have generated phenotypic data for 3,328 genes, providing unprecedented insight into how these genes function and their role in rare human diseases.

The considerable success of mouse models in providing insight into the biological mechanisms underlying human disease rests on the genetic and physiological similarities between the two species and the relative ease with which mice can be bred and housed in laboratories. However, despite their widespread use, mouse knockout experiments have been infamously difficult to reproduce, and standardized protocols for functional annotation have been largely lacking.

With this in mind, the IMPC has aimed to conduct standardized phenotypic screens on a large collection of mice carrying targeted null mutations in protein-coding genes. The consortium has already dedicated substantial effort to developing standardized phenotyping protocols and quality-control procedures to reduce unwanted technical variation across the ten participating centers². Their automated statistical analysis pipeline was designed to measure 509 phenotypes (encompassing neurological, behavioral, metabolic, cardiovascular, pulmonary, reproductive, respiratory, sensory, musculoskeletal and immunological traits) in 9- to 16-week-old knockout mice. The inbred C57BL/6N mouse strain was chosen for a uniform genetic background to decrease phenotypic variability.

Meehan *et al.*¹ now report the phenotypic characterization of 3,328 genes using this approach. Over half (1,830) of these genes had no existing mouse model, and 189 had no reported functional annotation. For 903 genes, the only available annotations had been inferred from computational analysis. The phenotypic data generated in this study provide some of

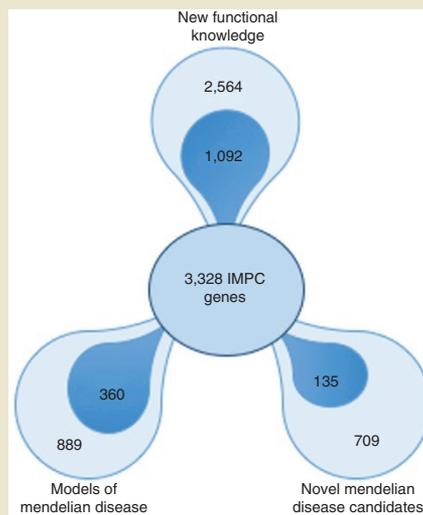


Image taken from Ref. 1

the first functional insights into a large number of previously uncharacterized genes.

Leveraging this enormous amount of new data, Meehan *et al.*¹ developed a computational pipeline, based on previous work³, that looks for similarities between the comprehensive set of phenotypes measured in the IMPC knockout mice and known clinical features of rare Mendelian disorders in humans. This analysis allowed the authors to identify 185 associations between human genes and diseases, the majority of which involve genes that have either not yet been studied in a mouse model or have not been reported as associated with a human disease. For example, they identified the first mouse models for diseases such as Bernard–Soulier syndrome type C, Bardet–Biedl syndrome-5 and Gordon Holmes syndrome, which may facilitate the discovery of new disease mechanisms and therapies.

Beyond providing new mouse models of human disease, the data set generated by the IMPC should be helpful in identifying the causal genes of specific diseases. “When you sequence individuals with rare diseases you often find a large number of disease-associated mutations but you’re not always sure what the causal gene is. And in many cases, the candidate genes have no known phenotypic information,” says Neal Copeland, Professor of Practice at the MD Anderson Cancer Center in Houston. “The IMPC data could be used to assign function to some of those mutations to help us

identify which mutations are actually causing the disease.”

The authors performed several analyses to validate their phenotypic pipeline using data from existing mouse model annotations. Of the 621 genes with mouse model data, only 385 had at least one phenotype in common with the IMPC phenotypes. 38% of previously reported gene–phenotype associations were also detected in this study. These results highlight the technical challenges facing studies of this kind. The authors speculated that the lack of reproducibility could be due to several factors, including different mouse genetic backgrounds and variations in experimental techniques and statistical methods.

By design, the new data set cannot shed light on certain aspects of human disease genetics. These include cases where the differences between human and mouse biology are substantial or where the mouse genetic background confounds the phenotypic measurements. In addition, Copeland says, “what’s being made here are null mutations, and many human diseases are caused by point mutations. So this is just the beginning.”

Even considering these limitations, the IMPC data set will continue to provide a valuable resource to the life sciences community. The broad scientific impact of the work, which has already led to >1,300 publications, can also be attributed in part to the open-access, user-friendly database set up by the consortium to distribute the data. As Copeland emphasizes, “One of the biggest challenges associated with a resource of this kind is creating a database that can disseminate information effectively across the world. An accessible database, which people outside of the mouse genetics field can also use, is key to maximizing the impact.” More generally, the IMPC’s systematic approach to capturing standardized phenotypic measurements across multiple centers, and the lessons learned along the way, will serve as an important example for researchers embarking on future large-scale, multinational collaborations.

Sarah Perry,
Associate Editor

1. Meehan, T.F. *et al. Nat. Genet.* **49**, 1231–1238 (2017).
2. de Angelis, M.H. *et al. Nat. Genet.* **47**, 969–978 (2015).
3. Smedley, D. *et al. Database (Oxford)* **2013**, bat025 (2013).