Submitter Name: Molly Bogue
Submitter email: Molly.Bogue@jax.org

New tools to assess research replicability using the Mouse Phenome Database

Molly A. Bogue[1], Gaurab Mukherjee[1], David O. Walton[1], Vivek M. Philip[1], Stephen C. Grubb[1], Matthew Dunn[1], Georgi I. Kolishovski[1], Jake Emerson[1], Timothy Stearns[1], Hao He[1], Vinita Sinha[1], Beena Kadakkuzha[1], Govind Kunde-Ramamoorthy[1], Elissa J. Chesler[1]

[1]The Jackson Laboratory

The Mouse Phenome Database (MPD; phenome.jax.org) is a widely used resource that provides access to primary experimental data, protocols and analysis tools for mouse phenotyping studies. We are curating data from inbred strains and other reproducible strains, including KOMP mice, Collaborative Cross (CC), CC-RIX, and founder strains. We are also collecting primary data from mapping populations, including advanced high-diversity mouse populations such as Diversity Outbred mice. New tools are under continual development and include replicability analysis. Even after rigorous standardization across laboratories, non-replicable results may be observed due to the interactions among genotypes and laboratories (*GxL*). An appropriate statistical model is a mixed model that ascribes a random effect to each laboratory, “Random Lab Model” (RLM). It therefore adds the interaction “noise”, σ^2_{GxL} on top of the individual animal noise to generate an adjusted “yardstick”, against which genotype differences are judged. Consequently, RLM raises the benchmark for declaring a significant genotype effect. σ^2_{GxL} cannot be estimated from a single-lab experiment, but it can be "imported" from previous multi-lab datasets. A *GxL* Replicability Adjuster tool has been built by our collaborators at Tel Aviv University (Yoav Benjamini and colleagues). This tool was recently implemented in the MPD system. A user selects existing MPD measures that are relevant to their data, uploads a per-animal datafile, and sets analysis options. A table is generated that shows unadjusted and *GxL*-adjusted results along with a comparison plot so that results can be easily visualized. This tool has been applied to addiction-relevant datasets from multi-lab studies.